

Towards the Biogeography of British Soil Microorganisms

Doctorate of Philosophy

Phillip James

April 2012

NEWCASTLE UNIVERSITY SCHOOL OF BIOLOGY

AND

CENTRE FOR ECOLOGY AND HYDROLOGY WALLINGFORD

Abstract

Soils are complex and highly variable ecosystems within which a multitude of diverse microbial populations can be found. Here the effects of co-correlating environmental variables and spatial separation upon the diversity and community structure of two microbial kingdoms is investigated using several molecular based community assessment methods. Using an optimised nucleic acid extraction procedure, suitable for landscape scale surveys of microbial biogeography, large scale bacterial and fungal targeted terminal restriction fragment length polymorphism (t-RFLP) analysis was undertaken on soils collected as part of the Countryside Survey 2007 to show that populations of both kingdoms are structured, in part, by variability in environmental and edaphic conditions as well as spatial separation. In the case of soil bacteria, pH was identified as the most important environmental variable, although computed models suggest that many other environmental variables also play strong roles. Fungal systems have weaker relationships with environmental variability and stronger spatial relationships, although dominant plant species and soil pH were shown to significantly affect community structure. However, comparing results generated from different genes with different taxonomic resolutions hinders accurate comparisons between divergent microbial kingdoms. Pyrosequencing analysis was undertaken on 15 geographically isolated soil samples forming a natural pH gradient to address the changes in bacterial and fungal populations at great sequencing depth and at taxonomic resolutions closer to the species level. Again, in both cases, members of these kingdoms responded to differences in soil pH and the above ground plant community. A number of bacterial and fungal taxa were found to be responsible for the changes in community structure and diversity noted in the t-RFLP based experiments and are proposed as candidates for indicators of soil pH. In summary soil bacterial and fungal populations are structured according to complex laws relating to co-correlating environmental and spatial variables. This leads to the hypothesis that microbial communities are structured in similar ways to terrestrial macro organisms, and thus ecological theories derived from observations of larger animals may be investigated within the microbial world.

Acknowledgements

I've always liked reading the acknowledgements in theses. They paint a picture of some triumphant, if slightly stressed and emotionally fragile PhD student, pouring out heartfelt gratitude to those who made the work possible. They seem to somehow bring a human edge to an otherwise emotionally vacant document, and I do not intend to disappoint in this regard. So, in a vain attempt to squeeze everyone who has helped me along the way, I will start with those without whom I would have been unable to write this.

Firstly, I would like to thank NERC and CEH for funding and allowing me to carry out the research which forms the basis of this thesis. I owe a great deal of thanks to the legions of people associated with the Countryside Survey who spent months collecting the soil samples, and without whom this study would not have been possible.

Dr. Rob Griffiths not only conceived the study but spent painstaking hours teaching me statistics and data analysis methods; put up with insane experimental suggestions; suffered monologues on the importance of microbial biodiversity; and read chapter after chapter of my thesis without complaining while providing useful and insightful comments without which this thesis would not have been possible. He also bought me a lot of beer.

Dr. Bruce Thomson basically taught me how to be a molecular microbial ecologist at the lab bench. His patience, direction, and willingness to teach were invaluable to a nervous and over eager first year PhD student. Throughout my PhD Bruce has been someone to bounce ideas off; read and correct my transfer reports; provide support when things went wrong; cast a critical eye over presentations; proof read chapters; and also buy me a lot of beer.

Prof. Mark Bailey and Prof. Andrew Whitley gave me the opportunity to study soil microbial biogeography as a PhD subject in the first place. Andy and Mark's enthusiasm for the project, and general support throughout, provided a source of constant confidence essential for any PhD student. I would particularly like to thank both for allowing me, and funding me, to present this work in Seattle. That was my first international conference and US experience, and was a definite highlight!

I would also like to thank the entire lab group and others at CEH Wallingford, particularly Dr. Anna Oliver, who provided much appreciated lab guidance and support, Dr. Christopher van de Gast for discussions on microbiology and "... the music you kids listen too", and Dr. Dan Reed for humouring my insane experimental ideas and useful discussions on the direction of my thesis. Tim Booth, Bela Telwari, and Bill Tyne all were essential in generating and analysing sequencing data and t-RFLP profiles through help with bioinformatics tools and maintenance of the DNA sequencer.

Support and understanding from family and friends is essential through a PhD so no acknowledgements would be complete without uttering the words "... I'd like to thank my Mum and my Dad..." in some form of style akin to a tear obscured Oscar awards ceremony. I can't put into words my thanks and do them justice so I will leave it at just one, thank-you (ok so it was two, but hyphenated). Rob James, Gary Brandon, Will Burge and Sam Shepherd, I'm sorry I never make it out on time to go fishing or to the pub, but thank you for sticking by me and still asking. Luke Breakspear, Frankie Meekings, Charlotte Robinson, and Barney Coleman would all be very annoyed if they didn't see their names here.

Finally I would like to thank my girlfriend Leonore Boelee, whom I dragged away from her native homeland of Holland only to experience the trials and tribulations of a manic postgraduate student in the final death throes of a PhD. Thank you for not complaining (too much!), the constant emotional support, and the food.... lots of food.

Table of Contents

Abstract.....	i
List of Figures	xi
Acronyms and Abbreviations	xiii
Chapter 1: General Introduction	1
1.1 The Soil Ecosystem	1
1.1.1 Soil has Functions.....	2
1.1.2 Soil as a Store.....	2
1.1.3 Soil Ecosystem Services.....	3
1.2 Soil Formation and Soil Diversity	4
1.2.1 Soil Forming Factors: Geology.....	4
1.2.2 Soil Forming Factors: Climate.....	4
1.2.3 Soil Forming Factors: Topography	5
1.2.4 Soil Forming Factors: Biology	5
1.2.5 Soil Forming Factors: Time	6
1.3 The Importance of pH in Modulating the Soil Ecosystem.....	6
1.3.1 As a Determinant of Bioavailability.....	6
1.3.2 Mediation of soil Function	7
1.3.3 Global Soil pH and Soil Acidification	7
1.3.4 Effect on the Soil Biosphere.....	8
1.4 Soil as a Habitat.....	8
1.4.1 The Soil Biota.....	9
1.5 Studying the Soil Microbiota.....	9
1.5.1 How to Study the Soil Microbiota: Molecular Approaches.....	10
1.5.2 Advances in Molecular Methods: Advantages and Disadvantages..	11
1.6 Current Knowledge of the Soil Microbiota: A Molecular Perspective	12
1.6.1 The Dominant Soil Microbial Biosphere	12
1.7 Soil Microbial Ecology and Biogeography	15
1.7.1 Traditional Ecology vs. Microbial Ecology	15
1.7.2 Patterns of Biodiversity: From Macro-bial to Microbial	16
1.7.3 Microbes: Everything is Everywhere... Or is it?	17
1.8 What Affects Soil Microbial Diversity?.....	17
1.9 Spatial Scale Affects Processes Shaping Diversity.....	18
1.9.1 The Microbial Scale.....	18
1.9.2 The Root Scale	19

1.9.3 The Field Scale	19
1.9.4 The Landscape Scale	20
1.10 Microbial Biogeography at Large Spatial Scales	21
1.10.1 Dispersal and Colonization	22
1.10.2 Environmental Influences upon Soil Microbial Biodiversity	22
1.10.3 Soil Bacterial Biogeography at the Landscape Scale: Current Knowledge	23
1.10.4 Soil Fungal Biogeography at the Landscape Scale: Current Knowledge	24
1.10.5 Spatial Autocorrelation within Microbial Populations	25
1.11 Thesis Aims	25
1.11.1 The Countryside Survey 2007	25
1.11.2 Thesis Overview.....	26
Chapter 2: Evaluation and Optimization of a Common Nucleic Acid Extraction Procedure for Large Scale Soil Microbial Biodiversity Studies	28
2.1 Introduction	28
2.1.1 Background.....	28
2.1.2 Aims	31
2.2 Materials and Methods.....	32
2.2.1 Sample Selection	32
2.2.2 Modified Nucleic Acid Extraction	34
2.2.3 Multiple Nucleic Acid Extraction	34
2.2.4 Nucleic Acid Quantification and Measures of Purity.....	35
2.2.5 PCR and T-RFLP	35
2.2.6 Statistical Analysis	36
2.3 Results	37
2.3.1 Evaluation of Nucleic Acid Yield from Clay Soils.....	37
2.3.2 Effect of Numerical Environmental Variables upon Nucleic Acid Yield and Purity.....	38
2.3.3 Effect of Soil and Aggregate Vegetation Classification on Nucleic Acid Yield and Purity.....	40
2.3.4 Generalized Linear Models to Calculate the Difference in Nucleic Acid Yield and Purity Explained by AVC and Soil type	41
2.3.5 Mapping of Nucleic Acid Yield and Purity across the UK	44
2.3.6 The Effect of Multiple Extractions on Soils Across a pH Gradient....	46
2.4 Discussion.....	50

2.4.1 Increased Nucleic Acid Yield from Clay Soils.....	50
2.4.2 Factors Affecting Nucleic Acid Yield and Purity	51
2.4.3 Effects of Multiple Extractions Across a pH Gradient	54
2.4.4 Conclusions	55
2.4.5 Further work.....	55
Chapter 3: Landscape Scale Terminal Restriction Fragment Length Polymorphism Analysis of Soil Bacterial Communities across the UK.....	56
3.1 Introduction	56
3.1.1 Background.....	56
3.1.2 Aims	58
3.2 Methods	59
3.2.1 Sample Selection and DNA Extraction.....	59
3.2.2 Polymerase Chain Reaction.....	59
3.2.4 Statistical Analysis	59
3.3 Results	62
3.3.1 Identification of t-RFLP Peaks using <i>in-silico</i> Digestion of a Global 16S rRNA Gene Database.....	67
3.3.2 Effects of Soil pH upon Bacterial Diversity	68
3.3.3 Multivariate Recursive Partitioning Analysis of Soil Bacterial Communities.....	73
3.4 Discussion.....	74
3.4.1 Effect of Soil pH on Bacterial Diversity.....	75
3.4.2 Environmental Influences on Soil Bacterial Community Structure ...	76
3.4.3 Mapping Bacterial Communities	78
3.4.4 Conclusions	79
3.4.5 Future Work	80
Chapter 4: Pyrosequencing Analysis of Soil Bacterial Communities over a Natural pH Gradient: A Comparison of Taxonomic Marker Genes.....	82
4.1. Introduction	82
4.1.1 Background.....	82
4.1.2 Aims	84
4.2. Methods	84
4.2.1. Sample Selection	84
4.2.2. Nucleic Acids Extraction Procedure	84
4.2.3 Massively Parallel Bacterial Tag Encoded FLX Amplicon Pyrosequencing (BTEFAP)	85

4.2.4. Sequence Processing and Data Analysis	86
4.3 Results	88
4.3.1 Sequence Processing	88
4.3.2 Effects of Environmental Variables upon Bacterial Community Dissimilarity	91
4.3.3 Dissimilarity of the V1 – V3 and V6 – V9 Regions of the 16S rRNA Gene	
4.3.4 Phylogenetic Analyses of Community Dissimilarity	93
4.3.5 Alpha Diversity	95
4.3.6 Differences in Community Composition between Variable Regions Targeted	96
4.3.7 Bacterial Community Composition Across a pH Gradient	97
4.3.8 Analysis of Dominant Taxa Abundance Closer to the Species Level	100
4.3.9 Shared OTUs Between and Within pH Groups	101
4.3.10 Indicator Taxa of Soil pH Groups	104
4.3.11 Other Notable Rarer Taxa	106
4.4 Discussion.....	107
4.4.1 Overarching Patterns of Soil Bacterial Community Structure with Environmental Variance	107
4.4.2 Variable Region Choice Affects Reported Bacterial Community Structure	109
4.4.3 Acidobacteria Show Strong Relationships with Soil pH.....	111
4.4.4 Other members of the Dominant Biosphere.....	112
4.4.5 Conclusions	113
4.4.6 Further work.....	114
Chapter 5: Large Scale Fungal Community Analysis from Soils across the UK.....	116
5.1 Introduction	116
5.1.1 Background.....	116
5.1.2 Aims	119
5.2 Methods	119
5.2.1 Sample Selection	119
5.2.2 Nucleic Acids Extraction.....	120
5.2.3 PCR Amplification of ITS Genes	120
5.2.4 T-RFLP	121

5.2.5 Clone Library Analysis	122
5.2.6 Statistical analysis.....	123
5.2.7 Clone Library Sequence Analysis	123
5.3 Results	124
5.3.1 Environmental Effects upon Fungal Diversity.....	124
5.3.2 Environmental Effects upon Fungal Community Structure.....	125
5.3.3 Clone Library Analysis	131
5.3.4 Changes in Community Composition	134
5.4 Discussion.....	138
5.4.1 Limitations to Fungal ITS t-RFLP	142
5.4.2 Clone Library Analysis	143
5.4.3 Conclusions	145
5.4.4 Further Work	146
Chapter 6: Pyrosequencing Analysis of Soil Fungal Communities	147
6.1 Introduction	147
6.1.1 Background.....	147
6.1.2 Aims	149
6.2 Materials and Methods.....	150
6.2.1 Sample Selection and Nucleic Acid Extraction.....	150
6.2.2 Massively Parallel Fungal Tag Encoded FLX Amplicon Pyrosequencing (F-TEFAP)	150
6.2.3 Sequence Processing	150
6.2.4 Data Analysis	150
6.3 Results	152
6.3.1 Sequence processing.....	152
6.3.2 Phylogenetic Estimates of Soil Fungal Community Dissimilarity	153
6.3.3 Alpha Diversity	155
6.3.4 Taxonomic Composition.....	156
6.3.5 Indicator Taxa Analysis	160
6.4 Discussion.....	166
6.4.1 Changes in Fungal Communities Across Environmental Gradients	166
6.4.2 Plant Effects on Fungal Communities	169
6.4.3 ITS and 18S rRNA Gene Comparisons.....	170
6.4.4 Global Differences in Dominant Fungal Lineages	170

6.4.5 Fungal OTU Sharing	171
6.4.6 Explanations for Changes in Fungal Composition Across Environmental Gradients	172
6.4.7 Conclusions	174
6.4.8 Future Work	175
Chapter 7: Investigating Spatial Autocorrelation in Soil Bacteria and Fungi	176
7.1 Introduction	176
7.1.1 Background.....	176
7.1.2 Aims	179
7.2 Materials and Methods.....	179
7.2.1 Sample Selection	179
7.2.2 DNA Extraction	180
7.2.3 T-RFLP Analysis	180
7.2.4 DGGE	180
7.2.5 Statistical Analyses	181
7.2.6 Overview of the Mantel Statistic.....	182
7.3 Results	182
7.3.1 Analysis of Bacterial Spatial Autocorrelation using t-RFLP	182
7.3.2 Bacterial Spatial Analysis of Two Distinct Soil Environments using DGGE	186
7.3.3 Analysis of Fungal Spatial Autocorrelation using t-RFLP	189
7.4 Discussion.....	192
7.4.1 Spatial Autocorrelation at Small and Large Geographic Distances	192
7.4.2 Community Variability and Taxonomic Resolution Affect Bacterial Spatial Autocorrelation	193
7.4.3 Spatial Autocorrelation in Fungal Populations.....	194
7.4.4 Comparisons with Macro organisms	195
7.4.5 What Can Cause Spatial Autocorrelation in Microbial Populations?	197
7.4.6 Does Functional Spatial Autocorrelation Exist?	198
7.4.7 Is It Really Spatial Autocorrelation?	198
7.4.8 Conclusions	199
7.4.9 Future Work	200
Chapter 8: Final Discussion	202
8.1 Overview of Thesis	202

8.1.1 Optimize a Soil Nucleic Acid Extraction Method	203
8.1.2 Broad Scale Patterns of Soil Bacterial Community Structure are Highly Related to Environmental Variability	204
8.1.3 Pyrosequencing Sheds more Light upon Changes in Bacterial Community Composition with Relation to Environmental Variability	205
8.1.4 Broad Scale Patterns of Fungal Biodiversity are not Similar to Those Observed in Bacterial Populations, but may be Describing Patterns at Different Taxonomic Resolutions.....	207
8.1.5 Fungal Pyrosequencing Confirms an Environmental Effect upon Community Assembly	209
8.1.6 Soil Bacteria and Fungi Show Evidence for Spatial Organisation, but Taxonomic Resolution Affects the Magnitude of this Phenomenon	211
8.2 Impacts upon Soil Microbial Ecological Theory.....	211
8.3 General Criticisms.....	216
8.4 Methodological Limitations.....	218
8.5 Future Work	220
8.6 Hypothetical Scenarios: The Perfect Study.....	223
8.7 Final Conclusions.....	223
Appendix A	226
Appendix B	227
Appendix B:.....	228
Appendix C	229
Appendix D	230
Appendix E	231
Appendix F.....	233
Appendix G	234
Appendix H	235
Appendix I.....	237
Appendix J	238
Appendix K	239
Appendix L.....	240
Appendix M.....	243
Appendix N	246
Appendix O	249
Appendix P	250
Appendix Q	251

Appendix R	252
Appendix S	253
References.....	254

List of Figures

Figure 1: Interactions between the Worlds Spheres.....	1
Figure 2: Soil Sampling Sites.....	32
Figure 3: Increased DNA Yield from Clay Soils.....	38
Figure 4: Variability of DNA Yield and Purity across Soil Type and Vegetation Classification.....	42
Figure 5: IDW Interpolation of DNA Yield and Purity across the UK.....	45
Figure 6: Additive DNA Concentrations for Multiple Extractions Across a PH Gradient.....	48
Figure 7: Principal Component Analysis of Bacterial Community Structure as Defined by t-RFLP from Single Extractions and 5 Pooled Extractions.....	49
Figure 8: Matrix plot of UK Wide Bacterial t-RFLP with Highlighted Important TRFs.....	62
Figure 9: PCA of Bacterial t-RFLP Profiles Highlighting Important TRFS.....	63
Figure 10: NMDS Ordination of Bacterial Community Dissimilarity.....	64
Figure 11: IDW Interpolation of Bacterial Community Structure across the UK.....	67
Figure 12: <i>In-silico</i> Digestion of Global 16S rRNA Database.....	68
Figure 13: Alpha Beta and Gamma Diversity of Bacteria Across a pH Gradient.....	69
Figure 14: Environmental Dissimilarity Decreases with Increasing pH.....	70
Figure 15: Relationship of Specific TRFs with Soil pH.....	71
Figure 16: IDW Interpolation of Specific TRFS across the UK.....	72
Figure 17: Multivariate Regression Partitioning of Bacterial t-RFLP Profiles.....	74
Figure 18: NMDS Ordination of Bacterial Communities by V1 – V3 16S rRNA Region Pyrosequencing.....	91
Figure 19: NMDS Ordination of Bacterial Communities by V6 – V9 16S rRNA Region Pyrosequencing.....	92
Figure 20: Sequence Dissimilarity of V1 – V3 and V6 – V9 Regions of the 16S rRNA Gene.....	93
Figure 21: PCoA of Unifrac Dissimilarity Matrices of the V1 – V3 and V6 – V9 16S rRNA Region Sequences.....	94
Figure 22: UPGMA Tree of V1 – V3 and V6 – V9 16S rRNA Region Sequences.....	94
Figure 23: Chao1 Rarefaction of Bacterial Pyrosequencing Data.....	96
Figure 24: Proportional Abundance of Bacterial Phyla from Pyrosequencing of Two Different Regions of the 16S rRNA Gene.....	97
Figure 25: Bacterial Community Composition Across a pH Gradient.....	98
Figure 26: Phylogenetic Tree of Bacterial V1 – V3 16S rRNA Sequences.....	103
Figure 27: Phylogenetic Tree of Bacterial V6 – V9 16S rRNA Sequences.....	104
Figure 28: Soils Used for Fungal t-RFLP Community Profiling.....	120
Figure 29: Primer Binding Sites for the ITS t-RFLP Analysis.....	121
Figure 30: Decrease in Fungal Beta Diversity as pH Increases.....	124
Figure 31: NMDS Ordination of Fungal t-RFLP Profiles.....	125
Figure 32: Diversity and Evenness of Fungal Population's Decreases Across the First NMDS Axis.....	128
Figure 33: Comparison of Bacterial and Fungal t-RFLP Profile Dissimilarity.....	130
Figure 34: Fungal Clone Library Diversity Estimates at Different Stringencies of OTU Clustering.....	132
Figure 35: PCA of Fungal Clone Libraries.....	133
Figure 36: Shared Fungal OTUs Between pH Groups.....	134
Figure 37: Phylum Level Assessment of Fungal Community Composition.....	135

Figure 38: Class Level Assessment of Fungal Community Composition.....	136
Figure 39: Bootstrapped Phylogenetic Tree of Fungal Clone Library Sequences.....	138
Figure 40: PCoA of Fungal Unifrac Based Dissimilarity Matrices.....	153
Figure 41: UPGMA Trees of Fungal Pyrosequencing Data.....	154
Figure 42: NMDS Ordination of Fungal Pyrosequencing Data.....	155
Figure 43: Chao1 Rarefaction of Fungal Pyrosequencing Data.....	156
Figure 44: Fungal Community Composition Across Soil Gradients by Pyrosequencing.....	157
Figure 45: Phylogenetic Tree of Ascomycota Lineages.....	162
Figure 46: Phylogenetic Tree of Basidiomycota Lineages.....	163
Figure 47: Phylogenetic Tree of Chytridiomycota Lineages.....	164
Figure 48: Shared Fungal OTUs Between Samples and Phylogenetic Placement.....	165
Figure 49: Sample Locations for DGGE Based Spatial Analysis of Bacterial Communities.....	180
Figure 50: Mantel Test of Bacterial Community Dissimilarity and Geographic Location.....	183
Figure 51: Correlation Between Environmental Dissimilarity and Geographic Distance.....	184
Figure 52: Correlation Between Bacterial Community Dissimilarity and Environmental Dissimilarity.....	185
Figure 53: Partial Mantel Correlogram Showing Bacterial Spatial Autocorrelation.....	186
Figure 54: Comparison of Bacterial Community Dissimilarity in <i>Lolium</i> and <i>Calluna</i> Dominated Soils.....	187
Figure 55: Mantel Test of Fungal Community Dissimilarity and Geographic Location.....	190
Figure 56: Correlation Between Environmental Dissimilarity and Geographic Distance.....	190
Figure 57: Correlation Between Fungal Community Dissimilarity and Environmental Dissimilarity.....	191
Figure 58: Partial Mantel Correlogram Showing Fungal Spatial Autocorrelation.....	191

Acronyms and Abbreviations

AIC.....	Akaike Information Criterion
AVC.....	Aggregate Vegetation Classification
BLAST.....	Basic Local Alignment Search Tool
Bp.....	Base Pair
bTEFAP.....	Bacterial Tag Encoded FLX Amplicon Pyrosequencing
C.....	Carbon
CEH.....	Centre for Ecology & Hydrology
CS [Survey].....	Countryside [Survey]
CTAB.....	Hexadecyltrimethylammonium bromide
DGGE.....	Denaturing Gradient Gel Electrophoresis
dsDNA.....	Double Stranded DNA
fTEFAP.....	Fungal Tag Encoded FLX Amplicon Pyrosequencing
GLM.....	General Linear Model
HA's.....	Humic Acids
HST.....	Honest Significance Test [Tukey's]
IDW.....	Inverse Distance Weighted
IQR.....	Inter quartile Range
ITOL.....	Interactive Tree of Life
ITS.....	Internally Transcribed Spacer [region]
Kb.....	Kilo base
Mb.....	Mega base
MUSCLE.....	Multiple Sequence Alignment by Log-Expectation
N.....	Nitrogen
NCBI.....	National Centre for Biotechnology Information
NMDS.....	Non Metric Multi-dimensional Scaling
OTU.....	Operational Taxonomic Unit
P.....	Phosphorus
PCA.....	Principal Component Analysis

PCoA.....	Principal Coordinate Analysis
PCR.....	Polymerase Chain Reaction
PEG.....	Poly Ethylene Glycol
PLFA.....	Phospholipid Fatty Acid [analysis]
QIIME.....	Quantitative Insights into Molecular Ecology
RAM.....	Random Access Memory
RDP.....	Ribosomal Database Project
RIFs.....	Rock Inhabiting Fungi
rRNA.....	Ribosomal RNA
S (i.e. 16S rRNA).....	Svedberg
SD.....	Standard Deviation
SE.....	Standard Error
SEM.....	Standard Error of the Mean
TRF.....	Terminal Restriction Fragment
t-RFLP.....	Terminal Restriction Fragment Length Polymorphism
UK.....	United Kingdom
UPGMA.....	Unweighted Pair Group Method of Arithmetic Mean
V [Region].....	Variable [Region]

Chapter 1: General Introduction

1.1 The Soil Ecosystem

“Soil is a natural body, differentiated into horizons of mineral and organic constituents, usually unconsolidated, of variable depth, which differs from the parent material below in morphology, physical properties and constitution, chemical properties and composition, and biological characteristics”

Hans Jenney 1941

Soil ecosystems are globally distributed and cover large proportions of the earth not covered by water (Buol *et al.* 2003). Composed of a mixture of air, water, rock and biological matter, soils can be said to be the principal interface between many of the world’s spheres, for example the atmosphere, hydrosphere, lithosphere and biosphere (Bohn *et al.* 2001) (Figure 1).

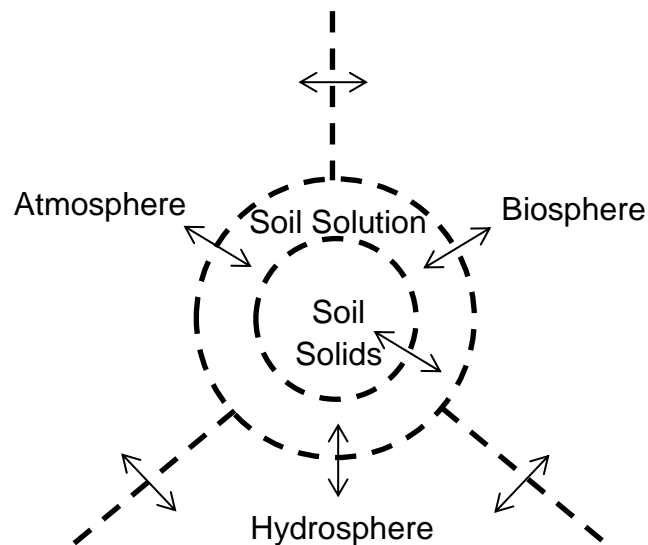


Figure 1: Interactions between the world’s spheres centres around the soil and soil solution. Figure adapted from Bohn *et al.* 2001 pg 2.

As an important constituent of the global ecosystem, the comparatively thin mantle of soil covering the earth (termed the pedosphere) is undoubtedly worthy of the many sub-disciplines which make up the study of soil science. Being a principal interface between the world’s spheres, many large scale functions are fulfilled by soil systems. However as well as mediating most global biogeochemical processes (Beare *et al.* 1995), simpler local scale functions, essential to the sustenance of the active biosphere as we know it today, are

also credited to the pedosphere (Mueller *et al.* 2010). These may be in the form of mediation of above ground financially important consumer driven processes such as agricultural practices (Hillel, 2009), or simply as a substrate for organisms to live in or on.

1.1.1 Soil has Functions

In terms of nutrient cycling, approximately 80 % - 90 % of primary productivity enters the soil through microbial decomposition of organic matter originating from dead plant litter and roots (Bardgett, 2005). Soil microbes are the principal decomposers of this organic matter. Fungi have the ability to produce a variety of extracellular compounds that allow the breakdown of recalcitrant cellular components such as lignocellulose, and hyphal networks allow the transport of these breakdown products to areas of lower nutrient availability (Carlile, 1995). Soil bacteria rely upon localised resources, such as carbon substrates, predominantly from plant detritus and root exudates (Bardgett, 2005). The solubilisation, mineralisation, and overall cycling of essential resources such as nitrogen, carbon, and phosphorus are largely controlled by the biological component of the soil and are of direct importance across multiple trophic levels. Approximately 97 % – 98 % of soil nitrogen is in the form of large macromolecules and complex insoluble polymers such as proteins, nucleic acids, and chitin, which are too large to be used by soil organisms directly (Heijden *et al.* 2008), and thus must be broken down into available soluble components via extracellular enzymatic activity. Nitrification and denitrification rates control the net production of biologically available nitrogen, and are almost entirely mediated by the soil microbiota.

1.1.2 Soil as a Store

Approximately 30 % of terrestrial carbon resides within the soil ecosystem, 75 % of which is in active exchange with the atmosphere making soils a potentially important sink for carbon released by fossil fuel combustion (Post *et al.* 1982). Estimates of carbon storage within the global soil ecosystem range from 700×10^{15} g (Bolin, 1970) to $2,946 \times 10^{15}$ g (Bohn, 1976) and therefore provide essential buffering capacity with relation to global CO₂ flux, but also a vast potential source of unreleased carbon. Recent anthropogenic manipulation of

soil systems has been shown to reduce the overall storage capability of soils (Schlesinger, 1977; Jenny, 1980) and thus large scale soil manipulation has a direct impact upon global carbon production and sequestration.

Solubilisation and mineralisation of phosphorus is controlled somewhat by microbial input through the release of phosphatases in low phosphate systems (Bardgett, 2005). A large proportion of terrestrial phosphate is stored in soil microbes (20 % – 30 %), greatly exceeding that of nitrogen (2 % – 10 %) and carbon (1 % – 2 %) (Bardgett, 2005). Labile phosphorus is comparatively low in soils, especially those of a low pH, due to fixation through adsorption to soil particles (Nwoke, 2003). Therefore microbial phosphorus is a large source of soil organic phosphorus (Brookes *et al.* 1984), and globally important for correct ecosystem functioning.

1.1.3 Soil Ecosystem Services

The cycling and production of nutrients essential for primary plant productivity can be banded together under the umbrella term “ecosystem services”.

“Ecosystem services” is generally regarded as a loose term for the products or functions provided or fulfilled by the ecosystem in question, for example water purification or carbon storage. A major ecosystem service supplied by soil is primary productivity in the form of agricultural practices. Global agricultural output is approximately \$ 1,500 billion per year (Source: Food and Agricultural Organisation of the USA, for 2010 data) and healthy, productive soils are reliant upon microbially mediated functions to fulfil this global requirement. With the world’s population exceeding 7 billion people (Source: World Bank, 2011), and rising rapidly, it is vital for a sustainable future to maintain healthy soils for food production, minimise anthropogenic climatic change, and maintain effective biogeochemical cycling in general (Banwart, 2011). The range of ecosystem services provided by soils is vast and may vary with each soil individually. It is the large biological, chemical, and physical diversity of soils found on the planet that allow these ecosystems to provide a vast array of services needed by human and non-human organisms alike for continued sustainable existence.

1.2 Soil Formation and Soil Diversity

In general, the chemical and structural diversity of soils is vast, and can vary considerably over small geographic distances (Ball & Williams, 1968). The formation of soil, pedogenesis, is the process by which this structural and chemical diversity is achieved. Pedogenesis is the product of a number of co-correlating factors which interrelate to produce this high spatial heterogeneity. There are thought to be 5 main soil forming factors, these being the parent material, climatic input, landscape and topography, biological interactions, and time (Jenny, 1941).

1.2.1 Soil Forming Factors: Geology

The underlying geology has a major influence upon the final physiochemical properties of soil (Jenny, 1941). The mineral composition of eroded rock varies greatly with the parent material from which it is derived. For example, base rich basalts have high calcium contents, while acidic rhyolites are high in silica and low in basic compounds (Bardgett, 2005). The parent material is also a major determinant of the grain size of the resultant soil, with high calcium containing rocks producing finer, positively charged particulates forming clay like soils with high moisture and cation retention capabilities (Christophereson, 1994). Harder parent materials such as granites produce coarser weathered products forming sandy soils which leach nutrients and have a lower water holding capacity (Jenny, 1941). The variation in the parent material is therefore not only a large determinant of soil physicochemical properties, but also characterises the composition of supported floristic communities.

1.2.2 Soil Forming Factors: Climate

Climatic variables regulate the degree of weathering that parent material undergoes (Brady & Weil, 1996), as well as determining other physicochemical properties such as soil pH (Thomas, 1996) and the vegetation supported by a particular soil type (Bakkenes *et al.* 2002). Temperature and levels of precipitation strongly govern the rates of chemical reactions and activity of the soil biota (Bardgett, 2005; Leiros *et al.* 1999) which, in turn, affect the overall physiochemical and biological composition of soils. Decomposition of organic material by the soil microbiota is highly dependent upon temperature, and it is

theorised that for every 10 °C increase microbial enzyme-catalysed reaction rates double up to approximately 30 °C – 35 °C (Bardgett, 2005). A reduction in temperature and increase in precipitation reduces the microbially mediated decomposition of organic matter forming soils such as peats (Wardle, 2002), which are important global carbon sinks.

1.2.3 Soil Forming Factors: Topography

Topography is also important in this regard as its influence upon soil development is largely a result of soil drainage and erosion rates. High lying soils tend to drain relatively freely due to gravity, whereas soils in low lying valleys are often waterlogged and can be largely anaerobic in nature or in extreme cases totally waterlogged, causing a decrease in microbial decomposition rates (Bardgett, 2005; Jenny 1941).

1.2.4 Soil Forming Factors: Biology

The biological component of the soil is vast (Giller, 1996; Fierer & Jackson, 2006) and has a large impact upon the overall physicochemical properties of fully formed soil. Both above ground and below ground communities are important in this regard, interacting to cycle nutrients and disrupt the soil matrix (Beare *et al.* 1995). Above ground communities affect the quantity and quality of organic material entering the soil, be this by leaf litter, general organic detritus, or root exudate compounds (Wardle, 2002). The below ground communities control the rate of decomposition and thus overall cycling of organic material (Bardgett, 2005). The ease of decomposition depends upon the quality of organic matter input, for example leaf litters with high levels of recalcitrant lignocellulose components have slower decomposition rates than easily palatable foliage produced by ecosystems such as deciduous forest (Cornelissen *et al.* 1999). Organic material entering the soil ecosystem in the form of root exudate compounds is used by plants to promote specific microbial growth in an effort to increase nutrient uptake through mutualistic interactions (Morgan *et al.* 2005). Such compounds affect not only the local physicochemical properties of soil (Nardi *et al.* 2000), but also the local microbial community structure (Walker *et al.* 2003). Coupled with this, disruption of the soil matrix by roots and other organic structures such as fungal hyphae increase the soil

surface area available for nutrient and gaseous exchange, and has a direct impact upon the rates of nutrient transport through space. Anthropogenic manipulation also affects the formation of soil and its resultant physicochemical properties, via disruption due to tillage practices (Franzluebbers *et al.* 1999), addition of fertilisers (Marinari *et al.* 2000), or exposure to pollution (Brady, 1996).

1.2.5 Soil Forming Factors: Time

All of the principal soil forming factors interact over geological timescales and thus a successional pattern to soil formation can be seen. Both the physical and chemical composition of a soil can change across time with younger soils exhibiting distinctly different patterns of biodiversity and physicochemical composition from older soils (Crocker & Major, 1955). Variability in the comparative effect of each soil forming factor across time result in a large diversity of soil types, each with different physicochemical and biological composition.

1.3 The Importance of pH in Modulating the Soil Ecosystem

The physicochemical and edaphic diversity of soils is large, resulting in numerous different soil types worldwide (Ball & Williams, 1968). Of particular interest to soil ecologists is soil pH due to its importance in determining factors such as the bioavailability of specific compounds (Barber, 1995; Rousk *et al.* 2009), as well as imposing intracellular stresses on soil organisms in the form of direct DNA damage, denaturation of essential macromolecules, inhibition of enzymatic functions, or alteration of membrane potentials (Jeong *et al.* 2008).

1.3.1 As a Determinant of Bioavailability

The solubility of particular compounds within the soil ecosystem is largely governed by soil pH and is often proposed as a mechanism to explain ecosystem scale patterns in biodiversity. For example Al^{3+} concentration and the extent of the labile phosphorus pool change with soil pH (Dalal *et al.* 2001; Bardgett, 2005). Al^{3+} is toxic to numerous soil organisms (Pietri & Brookes, 2008), but its bioavailability is often low due to the formation of insoluble complexes with clay minerals (Dalal *et al.* 2001). Under low pH conditions,

approximately between pH 4 and 5, the structure of these clay minerals becomes unstable and Al^{3+} is released in soluble forms, increasing its bioavailability to the soil microbiota (Pina & Cervantes 1996). Low pH systems often have reduced available phosphorus due to the formation of insoluble phosphates with aluminium and iron (Hinsinger, 2001). The solubility of these phosphates increases with soil pH, with the greatest rates of phosphate release arising between pH 6 and 7 as at high pH it is lost to other inaccessible forms through interactions with calcium (Hinsinger, 2001). The size of the labile phosphorus pool is therefore greatest across this pH range, often implying higher bioavailability of phosphorus to plants and other members of the soil biota than in soil of more extreme pH.

1.3.2 Mediation of soil Function

Microbially mediated ammonia and nitrate oxidation rates are generally considered to be optimal in neutral and slightly alkaline systems through observation of pure culture isolates (Allison & Prosser, 1993). The exponential decrease in ammonia availability with decreasing pH is thought to be a factor affecting ammonia oxidation rates, although examples of both ammonia and nitrate oxidation have been seen in soils of acidic and alkaline nature (Allison & Prosser, 1993; Boer *et al.* 1991). Global scale analysis of nitrification rates versus soil pH showed no significant linkage, suggesting local scale variability in soil pH or soil structure may also influence global rates of soil nitrification (Booth *et al.* 2005). However local scale surveys of nitrification rates and pH do show a significant link suggesting that across short environmental gradients nitrification processes occur at greater rates in neutral soils than in those with an extreme pH (Norton & Stark, 2011; De Boer & Kowalchuk, 2001).

1.3.3 Global Soil pH and Soil Acidification

The majority of soils worldwide are naturally acidic (Bardgett, 2005), with boreal forests, bogs, and heathland often having pH values less than 4 qualifying as extreme acid, or ultra-acidic soils, as determined by the United States Natural Resources Conservation Service. Factors which affect soil pH are numerous and can involve climatic impacts such as acid rain (Krug & Frink, 1983) or the leaching of base cations due to high precipitation rates (Thomas, 1996).

Anthropogenic manipulation of soil, in the form of fertiliser addition or mine spoilage deposition (Johnson & Hallberg, 2005), also locally reduces soil pH through the dissociation of H^+ ions via the solubilisation or microbial conversion of ionic compounds. Decomposition of organic matter containing high levels of phenolic and carboxyl groups liberates hydrogen ions, as does the microbial oxidation of ammonium to nitrate, lowering soil pH (Bardgett, 2005). Carbonic acid formed from atmospheric carbon dioxide can dissociate in water to liberate hydrogen ions, further reducing soil pH (Johnston, 2004), while soils formed from basic parent materials or those with buffering capacities often exhibit pH values closer to neutral (Jenny, 1941). Plants remove ionic nutrients from the soil, often in cationic form rather than anionic. As a neutral charge must be maintained within root systems to maintain correct movement of charged compounds through plant tissue, H^+ ions can be secreted by root tissues to compensate for the imbalance in cation/anion uptake. In some instances plants also use secreted H^+ ions to locally acidify soils in an effort to increase the solubility and thus uptake rate of nutrients such as Iron (De Vos *et al.* 1986).

1.3.4 Effect on the Soil Biosphere

Many organisms exhibit an optimum pH range for growth, for example microbes (Fierer & Jackson, 2006), plants (Islam *et al.* 1980), and other members of the soil meso and macro fauna (Bardgett, 2005). Extremophiles with acid or alkaline tolerance do exist and examples have been documented within extreme pH systems such as hot springs or soda lakes with pH values close to 1 or 11 respectively (Kristjansson & Hreggvidson, 1995). These include acidophilic microorganisms such as *Thiobacillus* sp. and members of the meso fauna, in particular enchytraeid worms (Cole *et al.* 2002), both found within extreme low pH systems. In general, organisms residing within ultra-alkaline conditions such as soda lakes are restricted to a reduced group of fungi and bacteria. These systems are comparatively rare globally.

1.4 Soil as a Habitat

The properties of fully formed soil are in no way simple in their origin, and the vast diversity of soil and the associated physicochemical properties are a result of the effect of each of the interacting spheres within the soil ecosystem

(Stockmann *et al.* 2011). As small changes in any one of the previously mentioned co-correlating soil forming factors may significantly affect the physicochemical properties of the fully formed soil, edaphic conditions across landscapes are highly heterogeneous due to local scale variability in soil forming factors. The resultant heterogeneous nature provides a huge array of niches available for numerous organisms of different kingdoms (Giller, 1996), resulting in one of the most biologically active and diverse systems on the planet (Decaens *et al.* 2006).

1.4.1 The Soil Biota

Essential global biogeochemical processes fulfilled by the pedosphere are reliant upon numerous interactions between all domains of life which reside within soil (Beare *et al.* 1995). Within soil there exists a huge diversity of organisms, ranging in size and complexity from the smallest viruses, through single celled bacteria, archaea, algae, protists and fungi, to more complex multicellular organisms such as nematodes and micro-arthropods. Larger terrestrial organisms (henceforth referred to as macro organisms) such as insects, mammals, and plants also interact with the soil ecosystem. Although all organisms within the pedosphere interact to drive ecosystem-scale processes, the microbial constituent is considered highly important (Torsvik & Ovreas, 2002) as all terrestrial biogeochemical processes are in some way reliant upon soil microbes (Nacke *et al.* 2011). Furthermore, the diverse range of microorganisms within soil is important for the production of many important products such as antibiotics, anti-cancer drugs, anti-fungal compounds, anti-parasitic agents, herbicides, insecticides, and growth promoters (Rondon *et al.* 1999).

1.5 Studying the Soil Microbiota

For years characterising microbial communities from the soil ecosystem has been largely restricted to a highly reduced snapshot of the community provided by culture methods. The majority of microbial cells within the soil ecosystem cannot be cultured under standard laboratory conditions, and between 0.1 % and 5 % of organisms identified through microscopy techniques grow in culture media at rates unsuitable for investigative purposes (Janssen *et al.* 2002), if at

all. Molecular based methods for assessing microbial community structure, diversity, composition, and function have come to the forefront of microbial community analysis, and are often phenotypic or genotypic in nature. Recently community assessment has become predominantly nucleic acid based (Ranjard *et al.* 2000), and revolves around determining the sequence variability in particular conserved genes. In order to examine the genetic variability between related taxa, a highly conserved taxonomically informative gene must be chosen and amplified by polymerase chain reaction (PCR), prior to analysis by a spectrum of molecular techniques.

1.5.1 How to Study the Soil Microbiota: Molecular Approaches

The 16S rRNA gene, 18S rRNA gene, and internally transcribed spacer (ITS) region have all been proposed as taxonomic markers for investigating the community structure and taxonomic placement of particular groups of the soil microbiota (Liu *et al.* 1997; Anderson *et al.* 2004; Lord *et al.* 2006). The 16S rRNA gene (in bacteria) and 18S rRNA gene (in fungi) are required for the production of ribosomal subunits and are highly conserved across lineages of each microbial kingdom. With the ability to accumulate mutations in regions which are not determinants of the ribosomal tertiary structure, predictable patterns of sequence variability (Neefs *et al.* 1990) can be observed via the use of molecular community profiling techniques such as terminal restriction fragment length polymorphism (t-RFLP) and denaturing gradient gel electrophoresis (DGGE). Direct sequence comparisons with databases of identified rRNA gene sequences can also provide best guess identification of organisms within complex microbial communities. Due to the coding nature of these genes, a limit to the sequence variability is imposed by the eventual formation of non-functioning gene products. Non coding regions such as the ITS region, used predominantly in fungal community profiling studies, allow a greater degree of sequence variability to arise between closely related taxa (Anderson & Cairney, 2004), and therefore discriminate at taxonomic resolutions closer to that of the species (Lord *et al.* 2006). However, this variability can cause problems during bioinformatic analysis and prevent the calculation of accurate multiple sequence alignments.

1.5.2 Advances in Molecular Methods: Advantages and Disadvantages

The molecular methods used to study soil microbial populations have numerous advantages and disadvantages due to the mechanisms by which they portray the community under study. Taxonomic resolution and applicability to high throughput studies are two key factors which must be taken into consideration when determining the optimal method with regard to the research question proposed. DGGE has a relatively high taxonomic resolution and it has been proposed that the method is sensitive enough to detect single base pair alterations in a gene fragment and thus discriminate between closely related microbial taxa reasonably well (Muyzer & Smalla 1998). However, technical difficulty often limits this theoretical high taxonomic resolution in diverse systems, and slight variations in gel formation prevent the accurate comparison of multiple gels (Fromin *et al.* 2002). Thus, DGGE is not suitable for use with high throughput studies examining multiple hundreds of samples simultaneously. T-RFLP is ideally suited in this regard as many different samples can be directly compared due to the inclusion of an internal size standard (Osborn *et al.* 2000). However, taxonomic resolution is sacrificed as multiple microbial lineages may produce restriction fragments of the same length and only base pair additions or deletions can be detected. Therefore t-RFLP can only provide an indication of general overarching patterns of microbial diversity across soil gradients at relatively low taxonomic resolutions (Fierer, 2007). Furthermore, these molecular methods are generally aimed at examining the overall community structure of a microbial population through the generation of a community “fingerprint”, rather than assigning taxonomic identifications to OTUs in order to examine community composition. Direct DNA sequencing of taxonomically informative genes not only allows the detailed interrogation of total sequence variability (as opposed to additions, deletions or bond strength as in t-RFLP and DGGE) but also allows high level taxonomic identifications to be made via the use of bioinformatic tools and curated sequence repositories. Traditional Sanger sequencing strategies, using a di-deoxy chain termination method, are utilised in parts of this thesis, but a preference has been given to using next generation sequencing. A major disadvantage to traditional Sanger sequencing approaches is the fact that clone libraries are usually quite small and thus only examine the dominant microorganisms. However, pyrosequencing produces multiple thousands of

DNA sequence reads per sample, and thus captures a larger proportion of the diversity present (Roesch *et al.* 2007). Although affected by problems such as higher error rates than Sanger sequencing chemistries (Kunin *et al.* 2010) and the large computational expense needed to analyse the vast amounts of data generated (Caporaso *et al.* 2010), pyrosequencing methods are rapidly gaining popularity within the field of environmental microbiology (Chistoserdova, 2010). All molecular methods used for studying complex microbial populations have advantages and disadvantages. However, studying these complex populations using a range of methods with different taxonomic resolutions is useful as it offers the advantage of studying the same organisms from different perspectives (Ramette & Tiedje 2006).

1.6 Current Knowledge of the Soil Microbiota: A Molecular Perspective

The estimated numbers of individual operational taxonomic units (OTUs) of microbial life within a single soil habitat are huge, ranging from several thousand to over a million (although the latter estimate is too large to be accurately verified with current methods) (Gans *et al.* 2005; Schloss & Handelsman, 2006). Soils are thought to contain a large proportion of the total genetic diversity seen on earth (Whitman *et al.*, 1998) and thus the microbial constituent of the pedosphere fulfil a wide range of metabolic processes (Torsvik & Ovreas, 2002). This high metabolic diversity is most obvious within the soil bacteria, as natural populations have the ability to degrade even the most recalcitrant xenobiotics including those toxic to most other organisms (Timmis & Pieper, 1999).

1.6.1 The Dominant Soil Microbial Biosphere

Soil bacteria and fungi are the most abundant groups within the soil micro-biota (Bardgett, 2005) however both have different evolutionary histories. Bacteria are prokaryotic and generally unicellular, relying upon either passive transport through the soil matrix or energy consuming flagellum-mediated movement. Identified as key organisms for organic matter decomposition and nutrient cycling, soil bacteria produce many of the extracellular enzymes needed to break down organic material and are required for sustainable maintenance of soil health and productivity. Bacterial phyla such as the Proteobacteria,

Acidobacteria, and Actinobacteria are ubiquitous in soils and collectively carry out a wide range of trophic functions. The Proteobacteria are Gram negative and often facultative or obligate anaerobes which can have chemoautotrophic or heterotrophic functions, and several divergent lineages have photosynthetic abilities (Stackebrandt *et al.* 1988). Within soils, the Alphaproteobacteria, Betaproteobacteria, Deltaproteobacteria, and Gammaproteobacteria make up large proportions of the total bacterial community (Spain *et al.* 2009), with the Alphaproteobacteria often dominating. *Nitrobacter*, in the Alphaproteobacteria, and *Nitrosomonas*, in the Betaproteobacteria, are free living chemoautotrophs responsible for the oxidation of ammonium to plant available nitrate. Rhizobia, also belonging to either the Alphaproteobacteria or Betaproteobacteria, fulfil this function in plant associated niches through the colonisation of root nodules in legumes. Found predominantly in low nitrogen soils (often due to waterlogging), leguminous plants provide plant derived carbon and protection from competition in exchange for biologically available nitrogen in a symbiotic mutualistic relationship.

The Actinobacteria are important in soils for humus formation (Berg & McClaugherty, 2008) and are generally regarded as good indicators of healthy soils due to their ability to produce a wide range of extracellular compounds needed to breakdown recalcitrant polymers such as cellulose and chitin (Van Dijk, 2008). The Actinobacteria are Gram positive and have a wide variety of morphologies and metabolic functions (Ventura *et al.* 2007). Their physical appearance can range from hyphal, resembling that of some soil fungi, through coccoid or rod-coccoid shapes (Ventura *et al.* 2007). Because of this, their taxonomic placement was uncertain, and until relatively recently they were classified within the fungal kingdom as the Actinomycetes (detailed in Stackebrandt & Schumann, 2006). Known for their secondary metabolite production, the Actinobacteria have a significant pharmacological potential and numerous antibiotic compounds such as actinomycin have been isolated from members of this lineage (Waksman & Woodruff, 1940).

The Acidobacteria are a comparably recent discovery, and the understanding of their importance in soils is currently emerging. High order taxonomies are at present elusive, but acidobacterial lineages can be grouped together into a number of phylogenetically divergent subdivisions (Kielak *et al.* 2008). The first

species discovery occurred in 1991 through the isolation of *Acidobacterium capsulatum* (Kishimoto *et al.* 1991) and a handful of further isolates have been cultured since this time, including species of *Holophaga* (Liesack *et al.* 1994) and *Geothrix* (Coates *et al.* 1999). The majority of acidobacterial taxa have not been cultured, and thus their ecology and metabolism are not well understood (Kielak *et al.* 2008). However, specific groups have been shown to have very high abundances in soils in terms of the total bacterial community (Janssen, 2006). Thus their importance within the soil ecosystem is an area of current study.

Soil fungi are eukaryotic and can form multicellular hyphal structures that permeate the soil substrate allowing cellular compounds to be transported throughout the organism (Bardgett, 2005). As key organisms within numerous decomposition and nutrient cycling pathways (Stark *et al.* 1972), a number of soil fungal organisms form symbiotic relationships with plants (Isaac, 1992). Approximately 95 % of vascular plants belong to genera that have mycorrhizal associations (Malloch *et al.* 1980). The exchange of plant-derived carbon and fungus-produced nutrients (such as labile phosphorus) allows increased growth rates of plants, and an environment for fungi comparably free of competition. There is also evidence to suggest that fungal plant symbionts provide associated plants with protection against soil borne pathogens (Gosling *et al.* 2006). Ectomycorrhizal fungi infiltrate inter cortical cellular space and are predominantly found in temperate and boreal forest ecosystems (Baxter & Dighton, 2005). There are thousands of individual species, primarily comprising members of the Basidiomycota (although some are members of the Ascomycota and Zygomycota), each often having specific relationships with particular plant species (Rinaldi *et al.* 2008). Arbuscular mycorrhizal fungi in the phylum Glomeromycota are of particular importance to agriculture as they infect the root cortical cells of numerous plants including grasses, herbs, and crops (Daniell *et al.* 2001; Gosling *et al.* 2006). They are also thought to have been a keystone group during the colonisation of land by plants of aquatic origins (Pirozynski & Malloch, 1975; Heckman *et al.* 2001). Mycorrhizal fungi which infect root cortical cells but do not form arbuscules are termed ericoid mycorrhizal fungi and are often found infecting plants such as dwarf shrubs in acidic, alpine, or tundra ecosystems (Gardes & Dahlberg, 1996). Ericoid

mycorrhizal fungi are typically ascomycetes, but many Ascomycota are not root associated, for example the Leotiomycetes. Non root-associated members of each phylum exist across the soilscape, many forming parasitic or pathogenic relationships with a variety of the soil biota.

1.7 Soil Microbial Ecology and Biogeography

As processes mediated by the pedosphere affect nearly all domains of life, exploring the mechanisms behind these processes is vital for a sustainable future for terrestrial life (Caron *et al.* 2009). The microbial constituent of the pedosphere, responsible for the majority of these processes, has been termed a “black box” of diversity (Tiedje *et al.* 1998) meaning that we can observe the processes occurring as a result of the chemical interactions within the soil microbial biosphere, however there are vast unknowns surrounding the organisms responsible for the ecosystem scale processes we see. There is limited insight into the actual organisms which fill the wide array of functional niches within soil, and very little is known about the overarching ecology of these organisms in general (Lauber *et al.* 2009). As the principal definition of ecology is the study of the relationships between organisms and their environment (Wyld & Partridge, 1975), in order to define a functional framework for the soil microbial ecosystem, it is important to have an ecological basis for it to fit onto. Simply, to understand the mechanisms by which these important processes occur, the ecology of these microbial populations must be studied.

1.7.1 Traditional Ecology vs. Microbial Ecology

The loose definition of ecology means that it encompasses a wide range of sub disciplines, each with divergent aims. However, all have the same core aim of trying to further the understanding of how specific ecosystems operate as a whole (Odum & Barrett, 2004). The concept of ecology as the study of organisms dates back to Aristotle (Keller & Golley, 2000), but it was not until the turn of the 19th century that it gained in popularity and was recognised as a scientific discipline. A pivotal cornerstone of ecological study was undoubtedly Charles Darwin's insights into the origin of species (Darwin, 1859). Observations based upon the distribution and community structure of animal and plant populations, by him and others, resulted in a cascade of subsequent

theories and hypotheses that have shaped modern ecology. The term given to the study of variation in abundance of organisms across spatial and environmental gradients is biogeography. Even though the cornerstone of modern ecology is an understanding of biogeography, the soil microbiota is difficult to study in this respect. This is predominantly due to the vast numbers of different microorganisms which inhabit the soil ecosystem, their small individual size (Hill *et al.* 2000), and predominantly non-culturable nature (Janssen *et al.* 2002). As a result, observations documenting the distribution and abundance of soil microbial populations with reference to environmental and spatial variability are largely absent. As this is the case, there is little ecological basis upon which to found more complex theories, as has occurred with macro organisms.

1.7.2 Patterns of Biodiversity: From Macro-bial to Microbial

In general, the patterns of species distribution across geographic areas are highly regular in relation to geographic gradients such as latitude and altitude (MacArthur, 1972; Diaz *et al.* 1998; Willig *et al.* 2003), along with the degree of habitat isolation and size (Bardgett, 2005). Historical landscape change processes such as glaciation and continental drift form a basis of spatial structure within terrestrial organisms, upon which ecological factors such as speciation, adaptation, competition, and extinction act to give rise to multi scale biogeographic patterns (Legendre & Fortin, 1989). Environmental gradients running across geographic space alter local conditions, allowing adaptation, speciation, competition and extinction to occur at different rates and with different outcomes producing the diversity of terrestrial organisms. Differences in earth histories between sites of similar ecological characteristics explain the variability in community structure observed within macro-organism communities, which are possible because of limitations to dispersal between populations (Martiny *et al.* 2006). However, it is uncertain if soil microbial communities conform to these principles in view of the small individual size and assumed ubiquitous dispersal capability of microbes (Finlay, 2002). As a result, landscape changing events which form physical barriers to macro-organism dispersal may not hinder the proliferation of micro-organisms across geographic space. If this is the case, processes by which microbial biodiversity is produced

and maintained rely solely upon environmental diversity, and are fundamentally different from those which govern the biogeography of macro organisms.

1.7.3 Microbes: Everything is Everywhere... Or is it?

The phrase “Alles is overall, maar het milieu selecteert” was coined by Baas Becking in 1934 and is translated in numerous environmental microbiology studies as “Everything is everywhere, but, the environment selects” (O'Malley, 2007). Baas Becking's statement was derived from observations made by Martinus Beijerinck of similar microorganisms from different environmental origins growing on a selective medium. He stated that “Germs [microorganisms] are cosmopolitan, and that their presence or absence can be predicted and practically produced by creating specific environmental conditions” (Wit & Bouvier, 2006). In direct contradiction to “Buffon's Law” of biogeography (reviewed in Cox & Moore, 2010) stating that “environmentally similar but geographically separated areas have different patterns of biodiversity”, Beijerinck validated his claims by arguing that the small individual size of microorganisms allows cosmopolitan dispersal, and highlighted his documentation of the dispersion of volcanic dust from Indonesia to Holland following the eruption of Krakatoa in 1883 (O'Malley, 2007). Although potentially flawed in its generation, it now provides a stable null hypothesis to be productively challenged.

1.8 What Affects Soil Microbial Diversity?

In general terms, a multitude of factors have been shown to affect the community composition of below ground microbial communities. Perturbation studies examining the effect of changes in the chemical and physical characteristics of a soil upon the below ground microbial community often result in a measured change in microbial community structure pre and post perturbation. Until recently the majority of microbial ecology studies have been based around the principle of manipulating a single or reduced number of variables to define the relationship between the manipulated variables and the subsequent change in microbial community structure and diversity. Although these studies show that below ground microbial communities respond to manipulations of edaphic conditions, none defines the holistic relationship

between patterns in soil microbial biodiversity and environmental conditions. Changes in soil pH (Fierer, *et al.* 2007; Rousk *et al.* 2010), soil nutrient status (Lauber *et al.* 2009), labile toxic compounds (Rousk *et al.* 2010; Wood *et al.* 1995), above ground plant communities (Gomes *et al.* 2003; Halling, 2001; Allen *et al.* 1995), soil structure (Lupwayi *et al.* 1998; Girvan *et al.* 2003; Lauber *et al.* 2008; Dequiedt *et al.* 2011), and land use history and practices (Lauber *et al.* 2008; Nacke *et al.* 2011) have all been shown to individually alter both the fungal and bacterial composition of the pedosphere in terms of community structure, diversity, or richness and evenness, but few studies have aimed to elucidate the relative importance of multiple environmental factors simultaneously.

1.9 Spatial Scale Affects Processes Shaping Diversity

In view of the small size of individual soil microorganisms in comparison to the soil meso and macro fauna, the spatial scale at which soil microbial biogeographic studies are conducted will ultimately affect the reported findings.

1.9.1 The Microbial Scale

At the microbial scale, the soil structure itself has a direct impact upon the spatial distribution of soil microorganisms, and this has been highlighted through studies examining the aggregation of microbes within the soil pore space (Ranjard, & Richaume, 2001). Within agricultural systems undergoing fertilisation treatments, approximately 80 % of bacterial individuals were located within soil aggregate micro-pore spaces, and a distinct difference in the bacterial community structure was seen between soils with large and small grain sizes. *Holophaga* and *Prostheobacter* species were seen to have higher proportional abundances within small grain soils, while Alphaproteobacteria dominated soil with larger grain sizes (Sessitsch *et al.* 2001). As this scale represents that of individual organisms, predation and parasitism of individual microbes ultimately affects the biogeography of the soil micro fauna, with fungicolous and bacteriocolous organisms producing a patchy distribution of soil microbial taxa over the μm to mm scale.

1.9.2 The Root Scale

At slightly larger spatial scales, for example within root associated soils, the above ground plant community directly affects the soil microbial community structure through interactions with roots and root exudate compounds (Baudoin *et al.* 2003; Haichar *et al.* 2008). Mycorrhizal fungi and root associated bacterial taxa, such as *Rhizobium* species, have been shown to have high abundances in root associated soils of particular floristic communities, primarily due to an increase in nutrient availability from root exudate compounds (Rouatt & Katznelson, 1961). Coupled with this, the inclusion of symbiotic organisms within plant root tissue provides shelter from predation by organisms such as nematodes or parasitic fungi, and from competition from other members of the soil microbiota (Stacy *et al.* 1991).

1.9.3 The Field Scale

At larger spatial scales, for example across a field, ecological gradients are larger and the effect of further biotic and abiotic influences upon the soil microbiota becomes apparent. For example, differences in the community composition of the supported floristic community may be more important for microbial biogeography at this scale than local predation or parasitism of the soil microbiota. Differences in the quantity and quality of organic matter input (e.g. leaf litter deposition) have been shown to be important in structuring the below ground microbial communities at this scale (Bardgett; 2005; Moore & Hunt, 1988), and soils supporting different crop types have been shown to harbour different microbial communities (Wieland *et al.* 2001). The rate of leaf litter decomposition, and thus input of carbon into the soil, can change with the above ground vegetation type and thus alter associated below ground communities (Wardle 2002). Leaf litters containing higher levels of recalcitrant plant polymers take longer to degrade and thus rates of carbon input into the soil are generally slower (Bardgett, 2005; Strickland *et al.* 2009). Specific organisms with the metabolic capabilities required to decompose recalcitrant plant polymers often proliferate in such instances, for example actinobacterial taxa or fungal lineages which produce extracellular enzymes able to break down substances such as lignin and cellulose. Manipulations of the above ground plant community at specific sites, via the long term addition of CO₂ (600 ppm over 5 years), have shown that fungal and bacterial community structure,

diversity, and richness are altered due to the effect of changing plant communities rather than the addition of CO₂ alone (Montealegre *et al.* 2002). Although the ecological gradients across these spatial scales are somewhat larger than those of soil micro habitats, they are smaller than those at the landscape scale. For example, changes in precipitation, soil pH, floristic communities and so on. are comparatively small across fine and field scales, and thus different factors will affect the community structure and diversity of soil microorganisms at larger spatial scales.

1.9.4 The Landscape Scale

Landscape scale assessments of soil microbial biodiversity aim to define the overarching relationships between the soil microbiota and the abiotic and biotic gradients found at this spatial scale. An entire landscape may encompass numerous soil types and habitats, each with fine scale differences in edaphic conditions. Therefore the relative effects of the processes which shape soil microbial communities will be different across an entire landscape when compared with local scale examinations. Biotic and abiotic gradients are large across an entire landscape compared with the local scale, and changes in abiotic and biotic characteristics of soils, such as pH, nutrient status, moisture content, climate, floristic communities, land use and levels of anthropogenic manipulation potentially become more important for defining microbial community structure and diversity than those highlighted as significant at the local scale. Deciding the correct spatial scale at which to examine microbial biodiversity depends upon the overall aim of the study in question. Landscape scale assessments of soil microbial biogeography offer the possibility to define the most basic relationships between external variables and the community composition of the soil microbiota, prior to highlighting local scale relationships within a given landscape. Ideally, global scale patterns in soil microbial biodiversity would be elucidated prior to mining down towards the landscape scale, then the field scale and so forth to define a multi scale assessment of soil microbial biodiversity. However, the limitations imposed by financial and technical constraints prevent this top down approach at present. Therefore, large scale studies, expressly examining the responses of microbial populations to variations in external factors at the landscape scale are essential for elucidating a general ecological framework for soil microbial communities, and

for furthering the understanding of how these diverse and important ecosystems operate at multiple spatial scales.

1.10 Microbial Biogeography at Large Spatial Scales

For a time, it was questioned whether a true biogeographic effect was occurring within soil microbial populations at all (Fierer & Jackson, 2006). Evidence suggests that certain microbial lineages have a cosmopolitan distribution within the global biosphere, and that therefore the landscape has no impact upon the distribution of these organisms in the environment (O'Malley, 2007). This is predominantly thought to be because large population sizes and short generation times result in high dispersal rates (Green & Bohannan, 2006). However, it is unquestioned that macro organisms with high dispersal rates, such as birds and butterflies, show strong biogeographic patterns across large spatial scales (Dolan, 2006). Soil bacteria are globally distributed (Lozupone & Knight, 2007) and thus, at this level of taxonomic resolution, the “everything is everywhere” hypothesis holds true. At the phylum level, certain taxa have been shown to have a worldwide distribution in both terrestrial and marine systems, for example the Betaproteobacteria, Cyanobacteria, Actinobacteria and Flavobacteria (Martiny, 2006). Moving closer towards the species level, specific genera of microorganisms have been shown to inhabit similar ecological niches that are separated by large geographic areas, for example specific hyperthermophiles in hydrothermal vents (Longnecker & Reysenbach, 2001). At the species level, cosmopolitan distributions frequently occur in human pathogens, such as *Escherichia coli* (Smith *et al.* 1991) and *Staphylococcus aureus* (Tristan *et al.* 2007). However, there is also a wealth of evidence that supports the existence of endemic taxa found within unique locations. For example, morphotypes of *Synechococcus* have been uniquely found in microbial mats within specific hot springs of North America (Papke *et al.* 2003), and certain *Pseudomonas* genotypes have been shown to be unique to isolated soils (Cho & Tiedje, 2000). Although specific cases highlight a cosmopolitan distribution of microbial taxa, and others imply an endemic nature, an often overlooked assumption is made that due to the apparent limitless dispersal inherent to organisms of this size, a true biogeographic pattern cannot exist within microbial populations (Fierer, 2008). This seems a somewhat misconceived notion as, if no biogeographic patterns are present within

microbial populations, then theoretically any sample taken from any biome would exhibit the same microbial community profile, and the relative abundance of all microbial taxa would be the same. A multitude of studies examining microbial community structure within the soil biome and others have shown this to be false (Buée *et al.* 2009; Dowd & Callaway, 2008; Fierer & Jackson, 2006; Rousk *et al.* 2010). This then leads one to assume that either an external influence is ordering and structuring these communities in relation to a set of ecological principles, or the community assembly of microbial populations is stochastic in nature and a high level of genome and metabolic plasticity allows specific lineages to exhibit a cosmopolitan distribution.

1.10.1 Dispersal and Colonization

From biogeographic studies of plants and animals, one of the key processes shaping macro ecological patterns is dispersal (Martiny, 2006). At present, there is some controversy surrounding the rate and extent of microbial dispersal, and even though endemism has been seen in certain microbial populations, meta-analyses of current literature have concluded that there is presently no basis for claiming that microbes have wide dispersal ranges using available data (Fierer, 2008; Jenkins *et al.* 2007). However, the direct assumption that the small individual size of a microorganism implies limitless dispersal seems to be a fallacy as the distance-mass relationship of passive dispersers is essentially random (Jenkins *et al.* 2007). Moreover, dispersal alone cannot affect biogeographic patterns unless successful colonisation of a habitat can be achieved. If colonisation rates are low, then endemism should conversely be high and this may be the case in the extreme environments previously mentioned, in particular hot spring microbial mats.

1.10.2 Environmental Influences upon Soil Microbial Biodiversity

Environmental selection is undoubtedly a key factor structuring the microbial communities which reside within soil ecosystems. Processes such as competition, predation, adaptation and other mechanisms which contribute to the overall structuring of populations within a particular ecosystem may only take effect if organisms which disperse to a new location find favourable conditions for proliferation. Therefore it is sensible to suggest that the most

important factor influencing the structure of microbial communities, and for that matter any population, is environmental selection. To elaborate, other mechanisms by which biodiversity is generated within a population may be important, but the primary bottleneck imposed upon individuals attempting to colonise new physical niches is that of the environmental conditions encountered by any organism when entering a new habitat. Of late, a focus within microbial biogeography has been to determine whether the environment does in fact play a large role in determining soil microbial community structure at the landscape scale, and if so, which environmental variables are the most important in this respect. As far as the soil ecosystem is concerned, surveys of microbial community structure and diversity have become popular as a means of determining overarching patterns of soil microbial biodiversity (Fierer & Jackson, 2006). Surveys of this nature are needed to simplify the study of the complex interactions within the vast diversity of soil microbial communities in order to gain an insight into the general ecology of these organisms. Such studies can be viewed as a “walk before you run” approach to determining biogeographic patterns; this seems to be an essential step in providing an ecological foundation upon which to base exploration of both more complex theories governing microbial biogeography and microbial ecology as a whole. Highlighting endemic and cosmopolitan taxa, or determining the community variability between areas of distinct environmental conditions, only indicates that a difference in microbial community structure and diversity is present. We know that the diversity of soil microbes is vast, and that they respond to a multitude of biotic and abiotic factors (Fierer & Jackson, 2006).

1.10.3 Soil Bacterial Biogeography at the Landscape Scale: Current Knowledge

The diversity and community structure of soil bacterial populations are hypothesised to be largely controlled by pH (Fierer & Jackson, 2006). However, due to the co-correlating nature of all environmental variables, the relative importance of other environmental variables is poorly understood. As highlighted previously, numerous studies have shown links between the community structure and diversity of different microbial populations and a number of environmental variables at different spatial scales. However, few studies have attempted a large scale holistic approach to defining the taxa

environment relationship within soil microbial communities at the landscape scale in order to determine the relative importance of each environmental variable concurrently. Where this has been performed, soil pH has been highlighted as an important determinant of bacterial diversity but different patterns relating bacterial diversity to soil pH have been proposed, including linear and unimodal distributions (Fierer & Jackson 2006). Furthermore different patterns have been proposed in different habitats, for example arctic soils (Chu *et al.* 2010). Patterns relating to the abundance individual bacterial taxa and pH have been highlighted, most notably within taxa belonging to the Acidobacterial subgroups (Sait *et al.* 2006).

Since some studies claim that soil pH is the sole determinant of bacterial community structure and diversity (Fierer & Jackson, 2006; Lauber *et al.* 2009), the mechanisms governing macro-bial and microbial community structure and diversity may be fundamentally different as macro-bial biogeography is determined by numerous co-correlating external variables rather than a one or a small number.

1.10.4 Soil Fungal Biogeography at the Landscape Scale: Current Knowledge

Studies examining the biogeography of soil fungal populations are less prevalent than for soil bacteria, and have yet to agree on a dominant driving force controlling the community assembly and maintenance of stable fungal communities. Smaller scale studies suggest that soil carbon content or nutrient status (Lauber *et al.* 2008) affects the distribution and community structure of soil fungal populations, whereas others suggest a definitive link between pH and fungal biomass production and growth rates (Rousk *et al.* 2009). Due to the interactions between numerous fungal phyla and specific plant species, the above ground vegetation type is hypothesised to be important in determining fungal distribution (Allen *et al.* 1995; Azaizeh *et al.* 1995; Grinstead *et al.* 1982). However, the distribution of plant communities also follows biogeographic laws and thus responds to environmental gradients and earth history events (Martiny *et al.* 2006).

1.10.5 Spatial Autocorrelation within Microbial Populations

Spatial effects upon soil fungal and bacterial populations have only recently become an area of interest; however different models suggest varying degrees of spatial autocorrelation within these kingdoms (King *et al.* 2010; Lilleskov *et al.* 2004; Nunan *et al.* 2002). As spatial relationships within the microbial world are becoming clear, the “everything is everywhere” hypothesis seems to be a reductionist view, and that the factors affecting soil microbial community structure may in fact not be entirely governed by the environment. If this is the case, then the biogeography of soil microbial populations may not be solely governed by environmental variability, and instead may conform to similar patterns to macro organisms, although potentially at different spatial scales. The combination of different large scale, high throughput examinations of the soil microbiota will increase insight into the factors which affect soil microbial community structure and diversity, with the eventual aim of forming a truly global and unifying model of soil microbial biodiversity.

1.11 Thesis Aims

The principal aim of this thesis is to investigate the effects of climatic, spatial, and edaphic conditions upon soil bacterial and fungal communities. This is achieved by using an array of nucleic acid based community profiling methods based upon high quality nucleic acid extracts from a wide variety of soil types. Furthermore, differences between the mechanisms which structure soil bacterial and fungal communities are discussed, along with the relative effects of spatial autocorrelation within the two kingdoms.

1.11.1 The Countryside Survey 2007

In order to perform this study a large number of soil samples was required, which spanned an entire landscape and represented a diverse range of habitats and physicochemical compositions. The Countryside Survey has been performed every 6 – 8 years since 1978, and provides scientifically reliable evidence about the state of the “health” of the UKs countryside and its relative change over time (Carey *et al.* 2008). As part of the full survey, over a thousand soil cores were obtained from every major habitat classification and it is these soil cores that provide the basis for this body of work examining the

biogeography of terrestrial microorganisms at the landscape scale. To provide a statistically representative set of soil cores for the UK, a stratified sampling design was conceived in 1978 and has been used in every subsequent survey since that time. Encompassing England, Scotland and Wales, 233 1 km by 1 km quadrants were chosen from geographically independent locations. From within each square, 5 randomly located 15 cm by 5 cm soil cores were taken as a stock for microbial community analysis, and 5 additional cores taken from adjacent locations for physicochemical analysis. For this study, sampling was carried out in June – August 2007 to represent the time of peak plant productivity. Each soil core was taken using a new sterile polypropylene coring device which was then sealed at both ends and acted as the storage vessel for each core. This was performed to avoid cross contamination between cores during the sampling and storage process. Soil cores were immediately stored at –20 °C prior to sample preparation and nucleic acid extraction. The top 15 cm of soil was used for microbial community analysis as this was thought to be the most biologically active part of the soil due to the close proximity to plant roots and the soil surface. Prior to microbial analysis, each soil core was split down its length under sterile conditions and homogenised by hand. No drying or sieving was performed as minimal anthropogenic manipulation of the soil core was desired prior to nucleic acid extraction. For nucleic acid extraction, 0.25 g of homogenised soil matter was aliquoted directly into the barcoded nucleic acid extraction bead mill lysis tubes. Not only does this survey provide a statistically representative set of soil samples for the entire UK, but the combination of these samples and detailed physicochemical and above ground measures of biodiversity provides a unique opportunity to address questions posed in relation to soil microbial biodiversity at an as yet unparalleled geographic resolution across an entire landscape.

1.11.2 Thesis Overview

After soil sampling and processing, the first step in this study was to develop an efficient nucleic acid extraction method suitable for use with a wide range of soil types in order to generate a national UK soil nucleic acid archive. The next step was to define the relationship between bacterial community structure and environmental variability at the landscape scale using community profiling methods such as t-RFLP and next generation sequencing. Next fungal

communities were assessed in a similar manner, and the effects of each environmental variable upon the biogeography of each kingdom were compared and contrasted. Finally the relationship between geographic space and community dissimilarity was examined within bacterial and fungal communities using t-RFLP and DGGE to determine the level of spatial autocorrelation at the landscape scale, and the relative effect of geographic isolation upon the community dissimilarity of the two microbial kingdoms under study.

Chapter 2: Evaluation and Optimization of a Common Nucleic Acid Extraction Procedure for Large Scale Soil Microbial Biodiversity Studies

2.1 Introduction

2.1.1 Background

The genetic diversity of soil microbes is immense (Torsvik *et al.* 1990) with estimates of bacterial species numbers ranging from a few thousand to potentially millions within a single gram of soil (Fierer *et al.* 2007; Gans *et al.* 2005; Schloss & Handelsman, 2006). To examine fundamental patterns in the distribution of soil microbial taxa, large scale surveys are required which encompass many thousand soil samples to represent landscape scale patterns in microbial biodiversity. Currently there are a number of molecular methods available to microbial ecologists that allow the analysis of complex microbial populations, such as terminal restriction fragment length polymorphism (t-RFLP) (Liu *et al.* 1997), denaturing gradient gel electrophoresis (DGGE) (Muyzer & Smalla, 1998), single stranded conformational polymorphism (SSCP) (Widjoatmodjo *et al.* 1994), DNA sequencing, and quantitative PCR. Each of these methods is polymerase chain reaction (PCR) based, and therefore dependent upon DNA of sufficient yield and quality as a starting point.

DNA of good yield and quality is essential to accurately examine microbial communities using nucleic acid based methods (Miller *et al.* 1999). Issues such as incomplete extraction, poor quality, or low yield of DNA will ultimately introduce artefacts into reported findings. Different soil types may have different extraction efficiencies and contain varying concentrations of PCR inhibiting compounds due to differences in physicochemical properties (Miller *et al.* 1999). Variation in the dominant microbial organisms present may also affect the microbial DNA extraction efficiency due to differences in cell wall construction (Pitcher *et al.* 1989).

Clay soils, in particular fine inorganic and fine organic clays, bind to nucleic acids through charged electrostatic interactions (Cai *et al.* 2007), Van der Waals forces, and hydrogen bonding (Taylor *et al.* 2002), thus reducing the overall yield of nucleic acids. Adsorption of DNA to clay particles occurs due to interactions between the charged groups on the nucleic acid macromolecule

and ionic compounds within the clay substrate. The fine nature of clay particles causes physical entrapment of the DNA molecules adding to the reduction in nucleic acid yield (Taylor *et al.* 2002).

Humic acids (HAs) are one of the most commonly co-extracted inhibitory compounds found in soils (Jizhong *et al.* 1996). They prevent Taq polymerase binding (Smalla *et al.* 1993) and interfere with restriction enzyme activity (Porteous & Armstrong, 1991). Concentrations of HAs can vary with soil type (Tebbe & Vahjen, 1993) with high organic content soils being typically problematic (Young *et al.* 1993). A number of post extraction methods have been proposed to remove HAs from crude nucleic acid extracts (Malik *et al.* 1994; Tsai & Olson, 1992; Ogram *et al.* 1987), but many of these involve multiple purification steps making them impractical for large scale, high throughput studies (Tebbe & Vahjen, 1993; Griffiths *et al.* 2000) and result in an inevitable loss of DNA yield (Roose-Amsaleg *et al.* 2001).

Differences between microbial organisms in terms of cell wall construction have the potential to prevent complete nucleic acid extraction through incomplete cell lysis (Zhou *et al.* 1996). Cell walls of Gram positive bacteria, such as the Actinobacteria, have high peptidoglycan content (Gontang *et al.* 2007), while fungal cell walls are fundamentally different in construction to prokaryotic organisms with multiple layers comprised of chitin, β -1,3-glucan, and mannoproteins (Lipke & Ovalle, 1998).

A number of different nucleic acid extraction protocols have been suggested for use with a wide range of soil types that have varying physicochemical properties. Many of these protocols are similar in nature, requiring some form of physical cell lysis stage (either in situ or post cell purification) coupled with a chemical precipitation step to obtain a crude nucleic acid extract (Tsai and Olson 1991; Picard *et al.* 1992; Griffiths *et al.* 2000). The high throughput nature of nucleic acid extraction procedures needed to perform studies examining multiple thousands of soil samples limits the use of extraction procedures to those that do not require multiple purification steps or long technical protocols (Volossiuk *et al.* 1995). Methods such as sodium dodecyl sulphate (SDS) extractions (Yeates *et al.* 1998), or those with chromatography (Picard *et al.* 1992), caesium chloride (Smalla *et al.* 1993), or gel purification (Young *et al.*

1993) steps produce high quality nucleic acid extracts, but are laborious to perform and require a large degree of technical expertise. Furthermore basic SDS extraction methods may not efficiently extract genetic material from Gram positive bacteria (Zhou *et al.* 1996). One method that is quick and gives reasonable results is the Griffiths method (Griffiths *et al.* 2000), which has been used in numerous DNA and RNA based assessments of soil microbial communities (Leininger *et al.* 2006; Manefield *et al.* 2002; Stach *et al.* 2001; Nicol *et al.* 2008; and cited by 497 other studies). Although comparatively quick and simple to perform, this method produces inconsistent results across multiple soil types and thus requires slight modification for use with large scale surveys examining multiple habitats and soil with diverse pedo-climatic and physicochemical properties. In particular, during preliminary testing of the Griffiths method, consistently low nucleic acid yields unsuitable for PCR amplification were obtained from high clay content soils. Coupled with this, to ensure that the method is suitable for use with large scale surveys examining multiple soil types simultaneously, the relative effects of pedo-climatic and physicochemical properties must be examined in relation to nucleic acid yield and purity.

A recent study of the factors affecting DNA yield from soils across France using an SDS extraction method highlighted several pedo-climatic variables as important determinants of DNA recovery, namely soil clay and fine silt content and to a lesser degree soil pH, carbon content, and nitrogen content. DNA yield was inversely correlated with sand and coarse silt contents as well as C/N ratio (Dequiedt *et al.* 2011). However, high organic content soils such as peats and peat topped soils are in a minority across France with the percentage coverage of these soil types falling below 0.6 % compared with values ranging from 10.9 % to 22.6 % in the UK (Montanarella *et al.* 2006). Furthermore, no indication of the effect of physicochemical properties upon DNA purity within the crude nucleic acid extract was attempted, although purification steps were proposed. Within the study, gel electrophoresis and ethidium bromide staining intensity were used to estimate DNA quantities in crude nucleic acid extracts (rather than total nucleic acid quantity). Although a valid approach due to the circumvention of biases introduced by soil impurities (Ranjard *et al.* 2003) the quantification method may not be as accurate as the analysis of multiple absorbance ratios

via spectrophotometry, which allows the quantification of co-extracted contaminants such as humic acids that may inflate reported DNA yields (Zhou *et al.* 1995). The study provides a useful insight into the factors which affect DNA yield across the French soil scape, but further experimentation is required to define the relationship between pedo-climatic variables, nucleic acid yield, and purity across a landmass with distinctly different soil composition and diversity using the method proposed here.

In addition to the above issues, it has also been recently reported that performing repeated extractions upon a single soil sample can affect the observed bacterial community structure (although not fungi), as well as increase total nucleic acid yield (Feinstein *et al.* 2009). Using pyrosequencing and commercially available soil DNA extraction kits, Feinstein *et al.* (2009) found higher proportions of Acidobacteria, Gemmatimonades and Verrucomicrobia in earlier extractions than later extractions, whereas the reverse was true for the Proteobacteria and Actinobacteria. As a number of these bacterial taxa have been highlighted as key organisms within high and low pH soils (Janssen *et al.* 2002; Jones *et al.* 2009), examining some of the potential artefacts introduced by the chosen nucleic acid extraction protocol is also critical to understand the potential error in results. Soil pH has been shown to be a key driver of bacterial community structure (Fierer & Jackson, 2006; Lauber *et al.* 2009), but the effect of soil pH on nucleic acid extraction efficiency using the Griffiths method has not been examined. Furthermore it is not clear if the effects of incomplete extraction highlighted by Feinstein *et al.* (2009) are responsible for the observed pH effect upon soil bacterial biodiversity, or if this is a legitimate phenomenon caused by selective pressures acting upon the soil bacterial population.

2.1.2 Aims

The aims of this chapter are to establish a rapid, reproducible nucleic acid extraction technique by modifying an existing method for use with multiple soil types found within the UK, and to establish correlations between pedo-climatic or soil physicochemical properties and nucleic acid yield and purity. In doing so, the generation of interpolated maps of nucleic acid yield and purity across the UK was attempted in order to relate patterns in nucleic acid yield and purity to a *priori* knowledge of soil conditions across the UK. Secondly, multiple sequential

extractions upon single soil samples spanning a pH gradient were performed to investigate the effect of incomplete extraction of nucleic acids upon bacterial community structure using t-RFLP, and to explore the effect of incomplete nucleic acid extraction in relation to the documented effect of soil pH on soil bacterial communities (Fierer & Jackson, 2006).

2.2 Materials and Methods

2.2.1 Sample Selection

Over 1000 soils from across the UK were sampled as part of the Countryside Survey 2007. A stratified sampling pattern was used to define 233 1 km by 1 km squares each representing one of 8 broad habitat classifications. From within each square, 5 replicate soil cores were taken randomly (Figure 2). For each 15 cm by 5 cm soil core, a range of edaphic and environmental variables was recorded, including soil pH, C:N ratio, moisture content, carbon content, nitrogen content, phosphorus content and loss on ignition, using a secondary core taken from an adjacent location. Climatic and biotic variables such as average annual sunshine, average annual rainfall, average cloud cover, the broad habitat classification, soil texture analysis, dominant plant cover, and the location of the sample site were also recorded. Aggregate vegetation classifications were used to define the general floristic composition and overall habitat type present at each sample site.

Clay soils were highlighted as exceptionally problematic during preliminary testing. Therefore, a subset of 22 high clay content soils (between 40 % and 100 % as defined by texture analysis) was selected for a comparative assessment using the published method (Griffiths *et al.* 2000) and the modified method described below (Appendix A). The subset was taken from independent sampling locations, with no two cores originating from any single 1 km by 1 km sampling site.

To assess the effect of multiple extractions upon the diversity and community structure of soil bacteria, a further subsample of 15 geographically isolated soils was selected. Bacterial diversity and the proportional abundance of the dominant bacterial biosphere species have been found to have specific relationships with soil pH (Fierer & Jackson 2006; Sait *et al.* 2006.) However, Feinstein *et al.* (2009) have shown that incomplete nucleic acid extraction has

specific effects on apparent proportional species abundance. Therefore, in order to determine whether the observed pH effect upon soil bacterial diversity and community structure was truly due to soil pH, or in fact due to incomplete extraction of nucleic acids, the soils used to analyse the effect of incomplete extraction spanned a natural pH gradient in replicates of 5 representing low pH (pH 4.3 \pm 0.23, n = 5), medium pH (pH 6.15 \pm 0.08 n = 5) and high pH (pH 8.28 \pm 0.16, n = 5). For sample locations and soil descriptions see Appendix B.

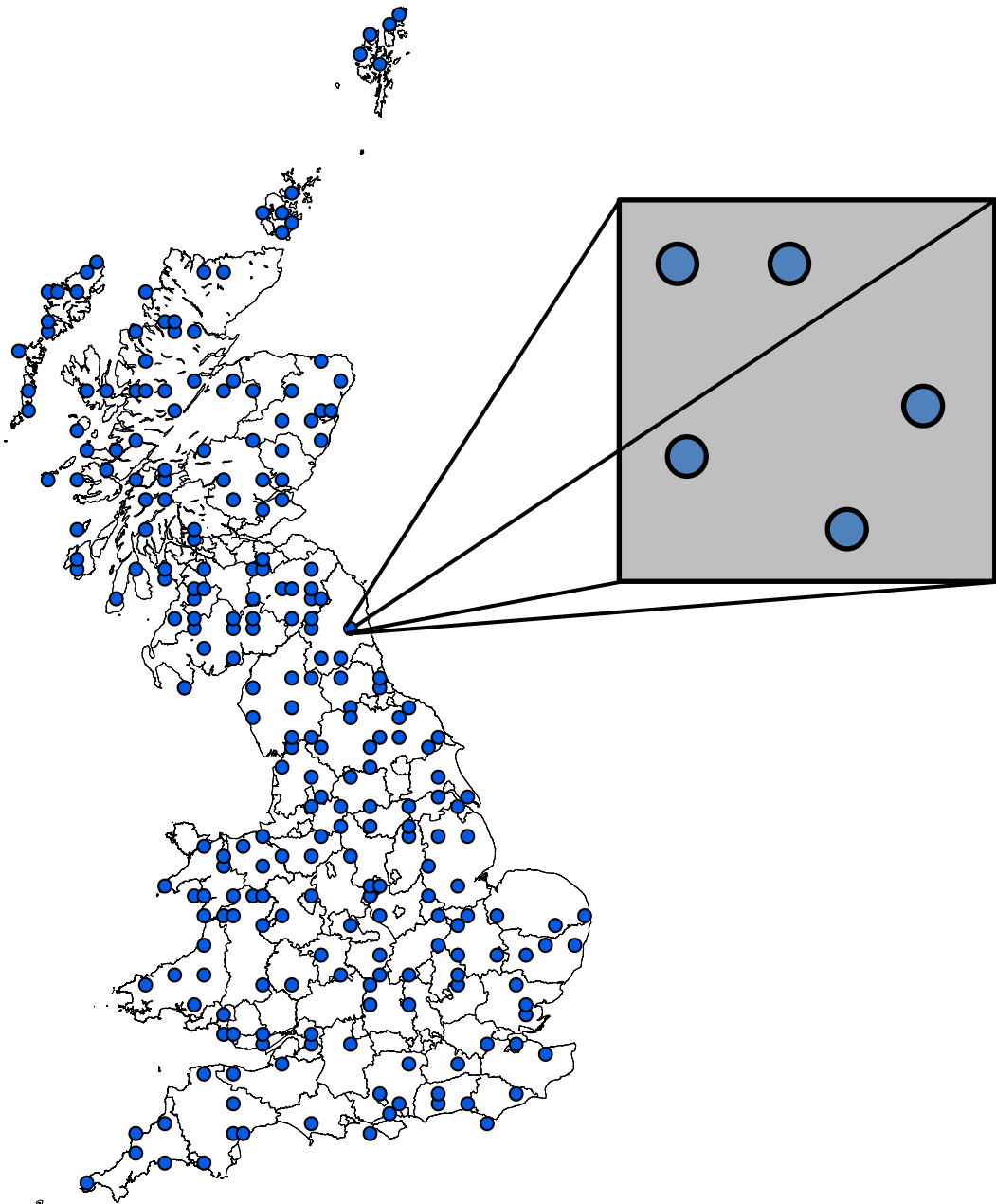


Figure 2: The 233 1 km by 1 km sampling sites distributed in an irregular sampling pattern to encompass every broad habitat classification within the UK. From each 1 km by 1 km sampling site, 5 15 cm by 15 cm soil cores were taken from randomly selected positions during the months of peak plant productivity (June – Aug) in 2007.

2.2.2 Modified Nucleic Acid Extraction

Based upon the Griffiths method (Griffiths *et al.* 2000), total nucleic acids were extracted from 0.25 g (wet weight) of homogenised soil in bead beating tubes containing lysing matrix E (MP Biomedicals, Illkirch, France) as opposed to 0.5 g as stated in the original method. To each sample, 0.5 ml of extraction buffer containing 10 % wt/vol hexadecyltrimethylammonium bromide (CTAB) with 240 mM potassium phosphate buffer (pH 8.0) in 0.7 M NaCl (Sigma, Poole, United Kingdom) was added. As a second modification to the method, tubes were then shaken by hand and frozen at -20 °C for approximately 1 h. Samples were thawed at room temperature and the procedure resumed as normal. 0.5 ml of phenol-chloroform-isoamyl alcohol (25:24:1 vol:vol:vol) (pH 8.0) (Sigma, Poole, United Kingdom) was added to each tube and cell lysis was achieved by bead beating in a FastPrep - 24 bead beating machine (MP Biomedicals, Illkirch, France) at 5.5 m s^{-1} for 30 s. The aqueous phase, containing nucleic acids, was then separated by centrifugation at $16,000 \times g$ for 5 min at 4 °C. An equal volume of chloroform:isoamyl alcohol (24:1 vol:vol) was added to remove phenol. Samples were then emulsified by shaking by hand, prior to a repeated centrifugation at $16,000 \times g$ for 5 min at 4 °C. Precipitation of nucleic acids was achieved by adding 2 volumes of 30 % (wt/vol) polyethylene glycol (PEG) 6000 (Fluka BioChemika, Buchs, Switzerland) in 1.6 M NaCl, overnight at 4 °C. Nucleic acids were then pelleted by centrifugation at $18,000 \times g$ for 10 min at 4 °C. Pellets were washed twice with 70 % (v/v) ice-cold ethanol then dried in a 5301 vacuum concentrator (Eppendorf, Cambridge, United Kingdom). Nucleic acids were dissolved in 100 µl molecular grade water (Sigma, Poole, United Kingdom).

2.2.3 Multiple Nucleic Acid Extraction

When multiple extractions were required to assess the impact of incomplete nucleic acid extraction upon bacterial community structure and diversity, extractions were performed five times for each sample as described above. All extractions were performed in triplicate. After the first round of extraction a further 0.5 ml of CTAB buffer was added to the retained soil/phenol mix, and the protocol resumed from that point. This included a further round of freezing and bead beating during each subsequent extraction.

2.2.4 Nucleic Acid Quantification and Measures of Purity

For every extraction, the nucleic acid yield was quantified using a Nanodrop 8000 spectrophotometer (Thermo Scientific, Wilmington, USA). The purity of each extract was defined by the 260 nm/230 nm and 260 nm/ 280 nm absorbance ratios. A 260 nm/ 230 nm absorbance ratio less than 1.8 indicates the presence of substances absorbing light at 230 nm (such as humic acids) while a 260 nm/ 280 nm absorbance ratio less than 1.8 indicates the presence of contaminants which absorb light at 280 nm (proteins and phenolics). Here, nucleic acid yield is reported as ng μl^{-1} instead of back calculated to per gram dry weight. This was done in order to avoid potential errors as nucleic acids were extracted directly from soils with minimal manipulation after sampling (i.e. drying and sieving).

2.2.5 PCR and T-RFLP

A ca. 446 base pair (bp) section of the V1 – V3 hypervariable region of the bacterial 16S rRNA gene was amplified using the primers 63f (5' – CAG GCC TAA CAC ATG CAA GTC – 3') (Marchesi *et al.* 1998) fluorescently labelled with 6FAM at the 3' end, and 519r (5' – GTA TTA CCG CGG CTG CCT G – 3') (Lane *et al.* 1991). PCR reactions were carried out using 1 μl of 1 in 10 dilutions of nucleic acid extracts with 0.2 μM of each primer, 5 U of Taq polymerase (Invitrogen, Paisley, UK), 1 mg ml^{-1} of BSA (New England BioLabs, USA), 0.5 mM MgCl_2 , and 15 mM dNTPs (Invitrogen, Paisley, UK) under the following conditions. A preliminary heating stage of 94°C for 90 s was used, followed by 29 cycles of denaturing at 94 °C for 45 s, annealing at 55 °C for 1 min and elongation at 72 °C for 90 s, and a final elongation step at 72 °C for 5 min. PCR products were purified using the Invitrogen PCR purification kit (Invitrogen, Paisley, United Kingdom) according to the manufacturer's instructions.

Subsequently, terminal restriction fragment length polymorphism (t-RFLP) analysis was carried out using a modification of a method previously described (Thomson *et al.* 2010). Purified PCR product (~50 ng) was digested at 37 °C for 4 h using 3 units of the restriction endonuclease Msp I (Promega, Wisconsin, USA), 1 μl of 10X enzyme buffer (Promega, Wisconsin, USA), and 0.1 μl of 10 mg ml^{-1} BSA. Total reaction volume was made up to 10 μl with ultra-pure, molecular grade water (Sigma Aldrich, St. Louis, USA). Fragment analysis was

carried out using an ABI 3730 DNA analyser (Applied Biosystems, California, USA) using the size standard Liz 600 (Applied Biosystems, California, USA). Binning of t-RFLP profiles was performed using the software package Genemarker v1.7 (Softgenetics, Pennsylvania, USA). Bins were defined manually, using a trace overlay of all samples. For each sample the relative abundance of terminal restriction fragments (TRFs) was calculated by dividing the fluorescence of individual peaks by the total integrated fluorescence of all peaks.

2.2.6 Statistical Analysis

All statistical analyses were performed using R (Ihaka & Gentleman, 1996) using the packages Vegan (Oksanen *et al.* 2011) and BiodiversityR (Kindt *et al.* 2005). Pearson's product moment correlation was used to measure the correlation between numeric environmental variables and nucleic acid yield and purity using the stats package in R (Ihaka & Gentleman, 1996). All environmental variables were scaled to have a mean of 0 and a standard deviation of 1 to directly compare variables with different units. Standard linear regressions were used to calculate the variability in nucleic acid yield and purity explained by environmental variables and standardised beta coefficients were calculated using the equation: -

$$\beta_{coeff} = coeff \times \frac{SD(x)}{SD(y)}$$

Where SD represents the standard deviation of the scaled environmental variable (x) or the response variable (y). The proportion of the variance explained by the whole model was partitioned into individual R^2 values for significant variables using the equation: -

$$R^2 = \beta_{coeff} \times \rho$$

Where ρ is the Pearson's correlation between the response variable (yield or purity) and the significant environmental variable.

The variation in nucleic acid yield and purity attributed to each categorical environmental variable (AVC and soil class) was calculated using generalized

linear models. Models were created using the `glm` function within the base package in R (Ihaka & Gentleman, 1996).

Where multiple linear regressions were performed, a Tukey's honestly significant difference post-hoc test was used to determine significant variables within ANOVA calculations in a pairwise fashion. Where this was performed, tables representing the pairwise comparisons are shown in appendices C and D.

Inverse distance weighted interpolation was performed to create maps of nucleic acid concentration and purity across the UK, using `ggplot2` (Wickham, 2009), `maptools` (Lewen-Koh & Bivand, 2011) and `gstat` (Pebesma, 2004) within R.

2.3 Results

2.3.1 Evaluation of Nucleic Acid Yield from Clay Soils

Due to the difficulty in extracting nucleic acids in sufficient quantities from clay soils using the published method (Griffiths *et al.* 2000), a direct comparison of nucleic acid yield obtained from clay soils by both the published and modified method, with a reduction of sample mass from 0.5 g to 0.25 g and an additional CTAB based freeze thaw step, was undertaken. A significant increase in nucleic acid yield from clay soils was observed when the modified method was used (ANOVA $p < 0.001$) (Figure 3). No significant differences in 260 nm/280 nm and 260 nm/ 230 nm ratios were observed (ANOVA $p > 0.1$). In a single case, no nucleic acid extract was obtained using either method; however, this was thought to be due to exceptionally high clay content.

Extracted Nucleic Acid Concentration for Original and Modified Method

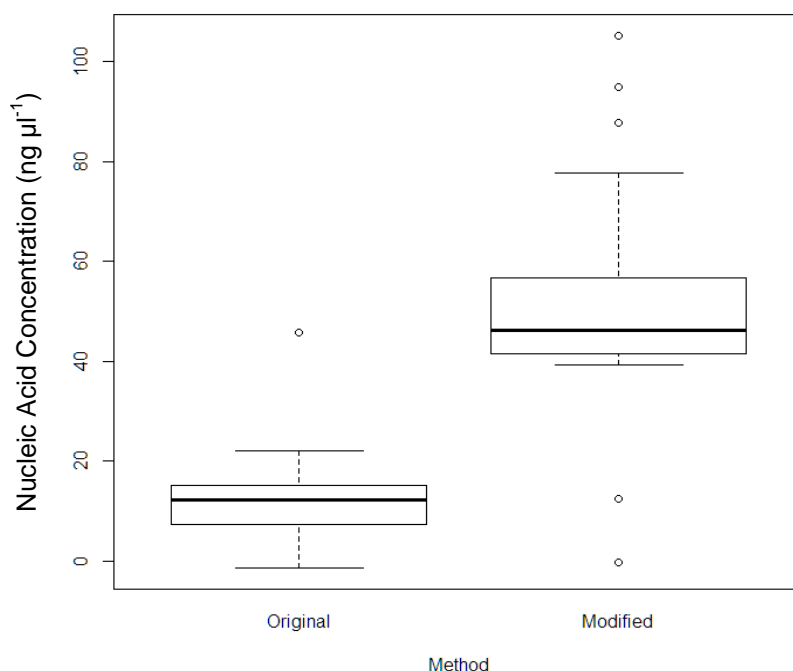


Figure 3: Box and whisker plot showing nucleic acid yield from 22 geographically isolated soils of high clay content using the original Griffiths method and the modified technique using a reduced sample mass of 0.25 g from 0.5 g, and an additional CTAB based freeze thaw step. Whiskers represent the range of the data after removal of statistical outliers, and points represent statistical outliers which are defined as > 3 times the inter-quartile range.

2.3.2 Effect of Numerical Environmental Variables upon Nucleic Acid Yield and Purity

Using the modified method, the mean nucleic acid concentration for all 1114 samples was $131 \text{ ng } \mu\text{l}^{-1}$ ($\text{SEM} = 5 \text{ ng } \mu\text{l}^{-1}$). Mean 260 nm/230 nm ratio was 1.72 ($\text{SEM} = 0.02$) and mean 260 nm/ 280 nm ratio was 1.83 ($\text{SEM} = 0.005$). Pearson's product moment correlation was used to measure the correlation between numeric environmental variables and nucleic acid yield and purity (Table 1). Scatterplots of variables with the strongest Pearson's correlations are shown in appendix E. Linear effects of physicochemical properties upon nucleic acid yield and 260 nm/ 280 nm ratios were minimal. However, moist acidic soils with high carbon contents had lower 260 nm/ 230 nm ratios, suggesting a slight increase in contamination by co-extracted compounds such as humic acids.

	Nucleic Acid Yield (ng μl^{-1})	260 nm/230 nm Ratio	260 nm /280 nm Ratio	Degrees of Freedom
Carbon (%)	-0.06 *	-0.34 ***	-0.11 ***	912
Nitrogen (%)	0.03	-0.29 ***	-0.08 *	912
C:N	-0.13 ***	-0.39 ***	-0.20 ***	912
Loss on Ignition (g)	-0.10 **	-0.32 ***	-0.07 *	993
Phosphorus (%)	-0.09 **	0.07	0.11 ***	954
pH	-0.11 **	0.41 ***	0.14 ***	1008
Moisture (%)	-0.02	-0.36 ***	-0.08 **	992
Altitude (m)	0.06	-0.22 *	-0.09	972
Average temperature	-0.02	0.30 **	0.11	978
Annual rainfall (mm)	0.07	-0.30 *	-0.08	978
Annual sunshine	-0.03	0.32 *	0.12 *	978

Table 1: Pearson's product moment correlation coefficients indicating levels of effect size between environmental variables and nucleic acid yield and purity. White represents no effect, light grey represents a weak effect and dark grey represents a medium effect as defined by Cohen (1988). No strong effects were observed for any numerical environmental variable. Significance values represented by asterisks where "****" < 0.001, "***" < 0.01 and "**" < 0.05.

Multiple regressions were used to compute standardised beta-coefficients and individual R^2 values for numerical environmental variables found to have a significant effect ($p < 0.05$) on either nucleic acid yield or purity (Table 2). Within the first multiple regression, modelling the relationship between environmental variables and nucleic acid yield, nitrogen content had a positive correlation, but accounted for a small proportion of the variance (1.8 %). Soil pH and loss on ignition both had weak but significant negative correlations with nucleic acid yield explaining 2.8 % and 6.3 % of the variance respectively.

Environmental Variable	Nucleic Acid yield (ng μl^{-1})			260 nm/230 nm ratio			260 nm/280 nm ratio		
	Beta	R^2	p	Beta	R^2	p	Beta	R^2	p
Nitrogen (%)	0.56	0.018	0.001	--	--	--	--	--	--
pH	-0.25	0.028	0.001	0.32	0.13	0.001	--	--	--
Loss on Ignition	-0.58	0.063	0.001	--	--	--	--	--	--
C:N	--	--	--	-0.28	0.11	0.001	-0.43	0.089	0.001
Total model R^2	0.13			0.19			0.06		

Table 2: Calculated beta coefficients and R^2 values for environmental variables found to be significant in a multiple regression analysis against nucleic acid yield, 260 nm/230 nm ratios, and 260 nm/280 nm ratios respectively. Although several environmental variables were found to have significant effects upon nucleic acid yield and purity, no variable explained more than 6 % of the variance in nucleic acid yield, or 13% and 8% of the variance in 260 nm/230 nm ratios or 260 nm/280 nm ratios respectively.

Decreasing soil pH and increasing soil organic matter content reduced nucleic acid yield marginally ($R^2 = 0.13$). However, the effect on the 260 nm/ 230 nm was greater, suggesting that low pH organic soils produce nucleic acid extracts with a higher concentration of co-extracted compounds which absorb light at 230 nm (i.e. humic acids) but only to a small degree ($R^2 = 0.19$). The effect on protein and phenolic contamination was minimal. However, very weak relationships with C:N ratio suggested that as soil organic matter content increases, contamination by substances absorbing light at 280 nm increases marginally ($R^2 = 0.06$).

2.3.3 Effect of Soil and Aggregate Vegetation Classification on Nucleic Acid Yield and Purity

As numerical environmental variables had only weak relationships with nucleic acid yield and purity, broad descriptions of soil type and habitat type were examined. Aggregate vegetation classification (AVC) and soil texture analysis were used to bin samples into groups defined by similar floristic or physicochemical properties.

The habitat type denoted by AVC had a significant effect upon nucleic acid yield (ANOVA: $p < 0.001$), as well as the 260 nm/230 nm ratio (ANOVA: $p < 0.001$) and 260 nm/280 nm ratio ($p < 0.001$). Tukey's honestly significant difference post-hoc test was used to compare mean nucleic acid yield, 260 nm/230 nm ratios, and 260 nm/280 nm ratios between each AVC in a pairwise fashion (Appendix C i-iii).

Again, only small differences in nucleic acid yield and purity were observed between the different habitat types denoted by the AVC (Figure 4a). Crops and weeds, heath and bog communities, tall grass and herbs, and lowland wooded areas resulted in significantly lower nucleic acid yields than upland wooded, fertile grassland, moorland grass mosaics and infertile grassland. Organic habitats such as moorland grass mosaics, heath and bog, and upland wooded areas returned the lowest 260nm/ 230 nm ratios, corroborating earlier findings suggesting that organic habitats produce nucleic acid extracts with a higher content of contaminants such as humic acids than improved fertile habitats such as improved grassland. The AVC had very little impact upon the 260 nm/ 280

nm ratios, suggesting that phenolic and protein contamination was minimal across all AVCs examined.

Soil type was also found to have a significant effect on nucleic acid yield and purity ($p < 0.001$). Clay soils returned the lowest nucleic acid yield and were significantly different from all soil types except clay-loam. Nucleic acid yields from silty-loam, peat, loam, and sand were not significantly different from one another and these soil types gave consistently the largest nucleic acid yields. Peat soils had significantly lower 260 nm/230 nm ratios than all other soil types, which were not significantly different from one another. Loam soils had a significantly higher 260 nm/ 280 nm ratio than all other soil types, which were not significantly different from one another (Figure 4b). Significant differences in mean nucleic acid yield, 260 nm/230 nm ratios and 260 nm/280 nm ratios between soil types are shown in Appendix D i-iii.

2.3.4 Generalised Linear Models to Calculate the Difference in Nucleic Acid Yield and Purity Explained by AVC and Soil Type

The effects of AVC and soil type on nucleic acid yield and purity were examined, both individually and combined, using GLMs in order to determine if these broad descriptors of habitat and soil type explained more variation in nucleic acid yield and purity than individual environmental variables. Calculation of the Akaike Information Criterion (AIC) showed that, in all cases, AVC and soil type should both be included within the model together, but the interaction term between the two was not significant. Combined, AVC and soil type explained 16.4 % of the variability in nucleic acid yield. Fertile grassland, infertile grassland, moorland grass mosaics, upland wooded, loam, peat and sand were found to be significant within the model, suggesting that these habitats and soil types had a significant effect upon the nucleic acid yield. For nucleic acid purity, again both AVC and soil type combined had significant effects (260 nm/230 nm ratio, $R^2 = 0.26$; 260 nm/280 nm ratio, $R^2 = 0.13$). For 260 nm/230 nm ratio, fertile grassland, heath and bog, infertile grassland, moorland grass mosaics, upland wooded, loam, and sand were significant in the model. As for 260 nm/ 280 nm ratios, all the above AVCs and soil classes were significant as well as the soil class peat. This shows that compared to other environmental variables,

both numeric and categorical, AVC and soil type combined were important determinants of nucleic acid yield and purity.

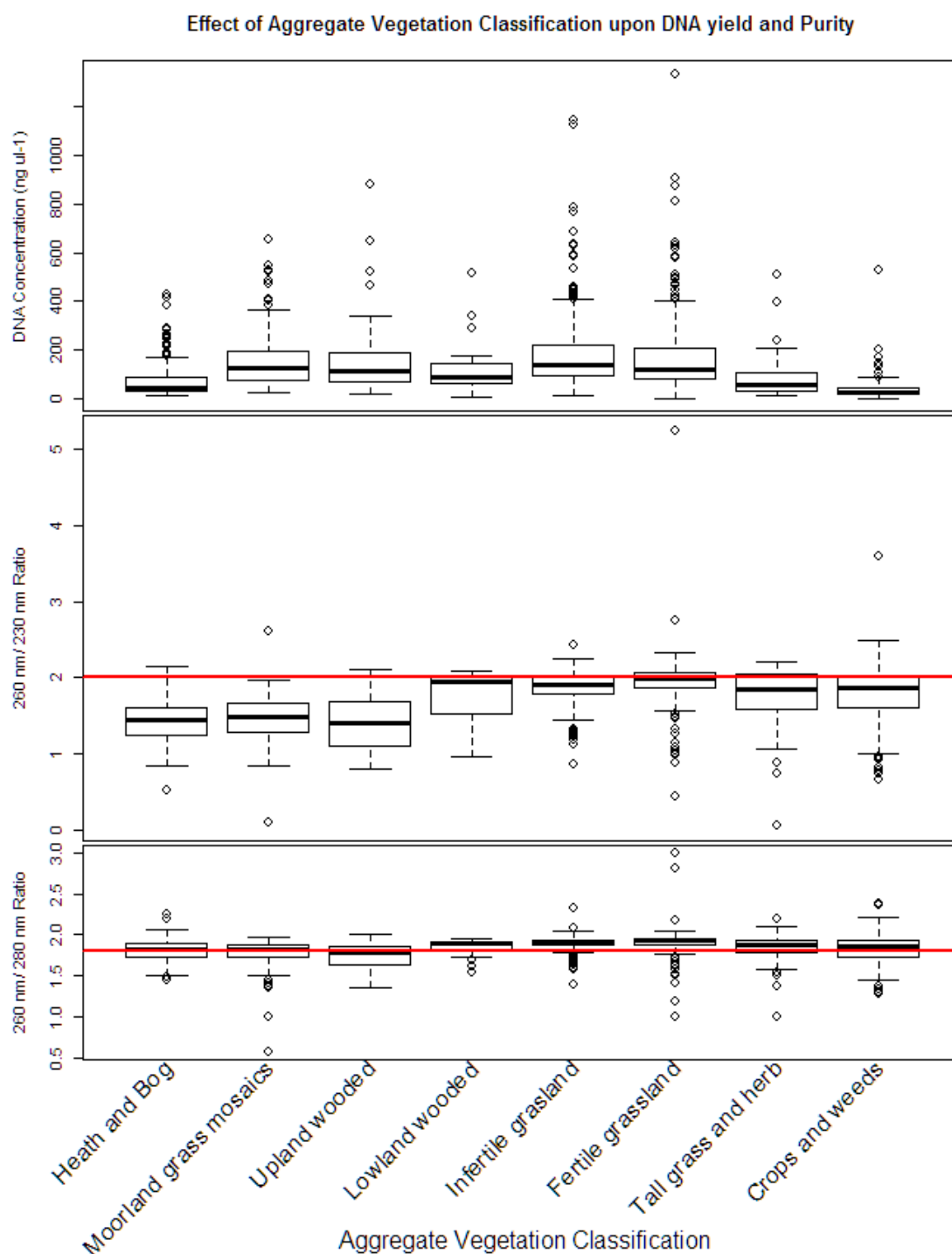


Figure 4a: Variability of mean extracted nucleic acid concentration, 260 nm/ 230 nm ratios and 260 nm/ 280 nm ratios as a result of differences in habitats defined by aggregate vegetation classification. Bold black lines represent the median of each data set, while boxes represent the inter quartile range (IQR). Error bars denote the range of the data after the removal of statistical outliers while points represent statistical outliers as defined by > 3 times the IQR. Horizontal red lines show desired 260 nm/ 230 nm and 260 nm/ 280 nm ratios.

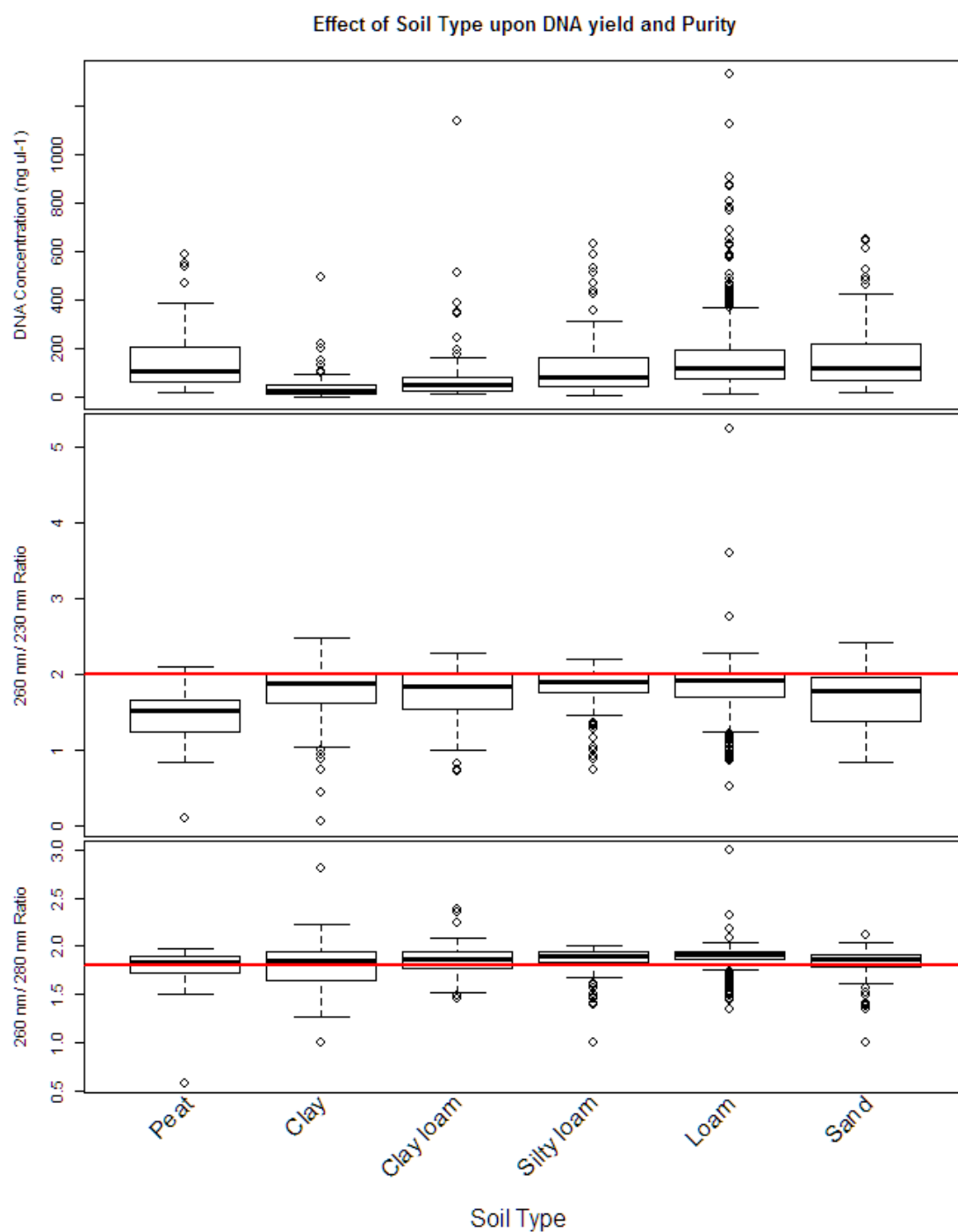


Figure 4b: Variability of mean extracted nucleic acid concentration, 260 nm/ 230 nm ratios and 260 nm/ 280 nm ratios as a result of differences in soil type. Bold black lines represent the median of each data set, while boxes represent the inter quartile range (IQR). Error bars denote the range of the data after the removal of statistical outliers while points represent statistical outliers as defined by > 3 times the IQR. Horizontal red lines show desired 260 nm/ 230 nm and 260 nm/ 280 nm ratios.

2.3.5 Mapping of Nucleic Acid Yield and Purity across the UK

Inverse distance weighted interpolation (IDW) was used to identify specific geographic areas which displayed an increase or decrease in nucleic acid yield or purity (Figure 5) and use *a priori* knowledge to try to explain these findings. Producing maps such as this is useful to highlight general patterns in soil nucleic acid yield and purity.

In terms of nucleic acid yield, most samples produced a yield between 100 and 200 ng μl^{-1} but a few samples located in areas of north Wales produced higher concentrations. The 260 nm/280 nm ratios were reasonably consistent across all geographic locations and, as shown in Figure 2, exhibited typical ratios close to the ideal value of 1.8. The 260 nm/230 nm ratios are generally higher than the 260 nm/280 nm ratios and causally a link can be made between higher 260 nm/230 nm ratios and higher concentrations of nucleic acid. This is most notable around the areas of north Wales highlighted earlier. The 260 nm/230 nm ratios are a good indicator of contaminants such as humic acids. Lower 260 nm/230 nm ratios are generally reported in samples from locations known to have large areas of organic soils such as peat land in north west Scotland and Wales. Mapping of 260 nm/ 230 nm and 260 nm/ 280 nm absorbance ratios shows little overall spatial variability in nucleic acid purity across the UK, suggesting that relatively pure nucleic acid extracts can be obtained from a wide range of spatially and physicochemically distinct soil types.

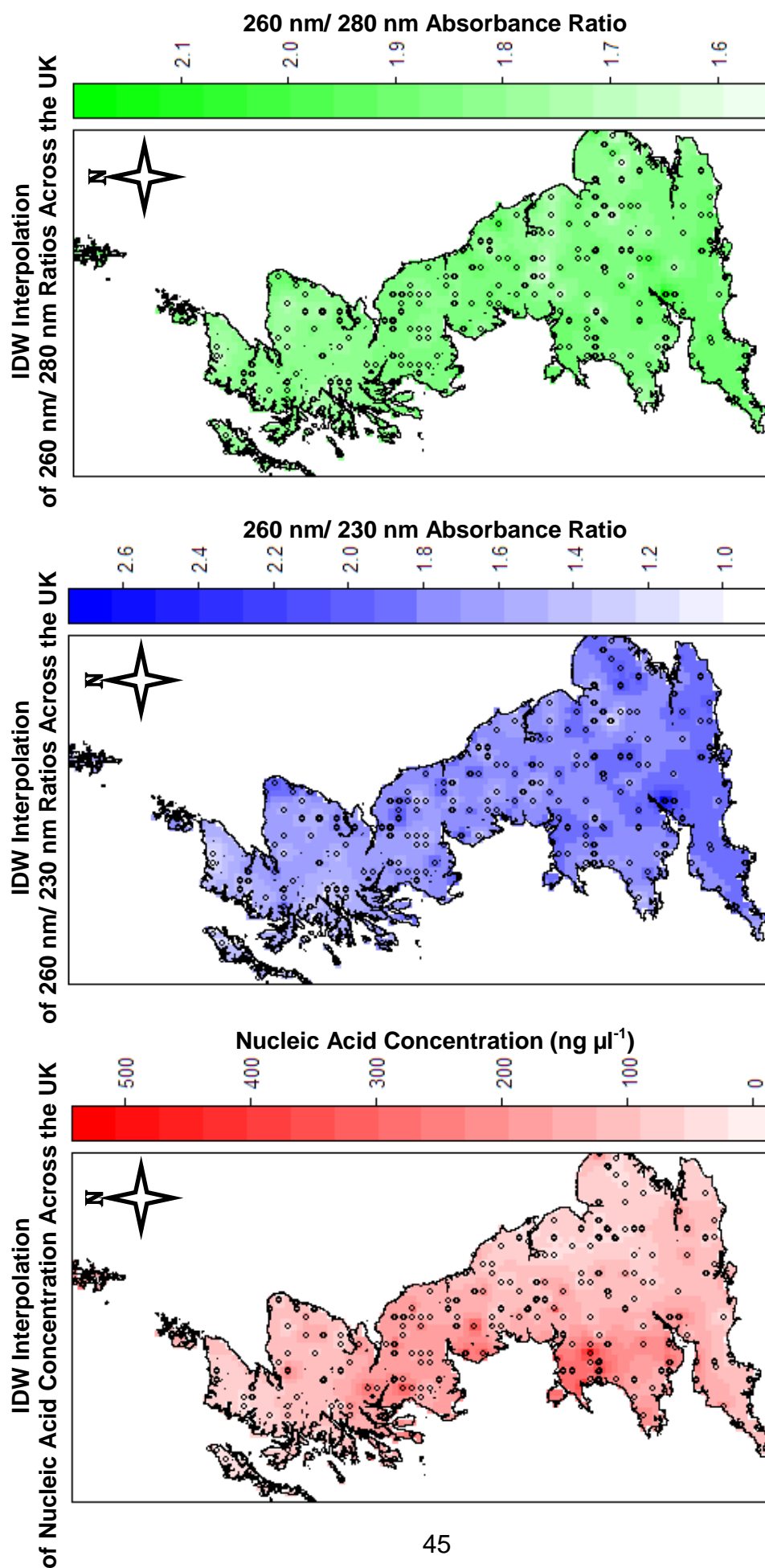


Figure 5: IDW interpolation of nucleic acid concentration (red) 260 nm/230 nm ratios (blue) and 260 nm/280 nm ratios (green) across the UK.

2.3.6 The Effect of Multiple Extractions on Soils Across a pH Gradient

Incomplete nucleic acid extraction has been highlighted as a potential problem within molecular analyses of bacterial populations from soil (Feinstein *et al.* 2009). Therefore the effect of multiple nucleic acid extractions upon single soil samples was examined. Furthermore, as soil pH has been highlighted as a dominant factor affecting soil bacterial diversity and community composition (Fierer & Jackson, 2006), the soils examined were chosen to span a natural pH gradient. For each sample, total nucleic acid yield increased with each extraction; however a plateau began to form after the third extraction (Figure 6).

To highlight any changes in bacterial community composition between the multiple extractions, principal component analysis (PCA) was undertaken. PCA showed that the effect of pooling multiple extractions on overall bacterial community structure was negligible compared with the primary extractions (Figure 7). Replicates of each extraction clustered together exemplifying a high degree of reproducibility. In all but one case, the bacterial community from the first round of extractions clustered with the communities from the pooled 5 extractions. This shows that, on the whole, multiple nucleic acid extractions had a negligible effect upon t-RFLP based estimates of bacterial community structure.

In one sample (Figure 7, low pH sample 5, red diamonds) a shift can be seen in ordination position of all replicates from the primary and pooled extractions. The hollow red diamonds representing the bacterial community structure elucidated from t-RFLP analysis of the primary extraction can be seen towards the centre of the plot in triplicate replication. Samples separate out on the basis of pH and in all but one sample the pooled extractions cluster with their respective primary extraction. Replicates cluster with one another highlighting the reproducibility of the nucleic acid extraction method. In the single sample where a change in community structure between the primary and pooled extractions is observed (low pH sample 5, all replicates), the pooled extractions still describe a community structure typical of low pH soils. However, the solid red diamonds representing the bacterial community structure from the 5 pooled extractions can be found towards the right of the ordination, again in triplicate replication. This suggests that, in this one sample, incomplete extraction had an effect on the reported bacterial community structure. TRF 52 is characteristic of the low

pH samples and, in the singular case highlighted above, an increase in the proportional abundance of TRF 52 can be seen due to repeated extractions. Therefore, repeated extractions do not contradict the overall pH effect upon bacterial community structure observed within the rest of the samples.

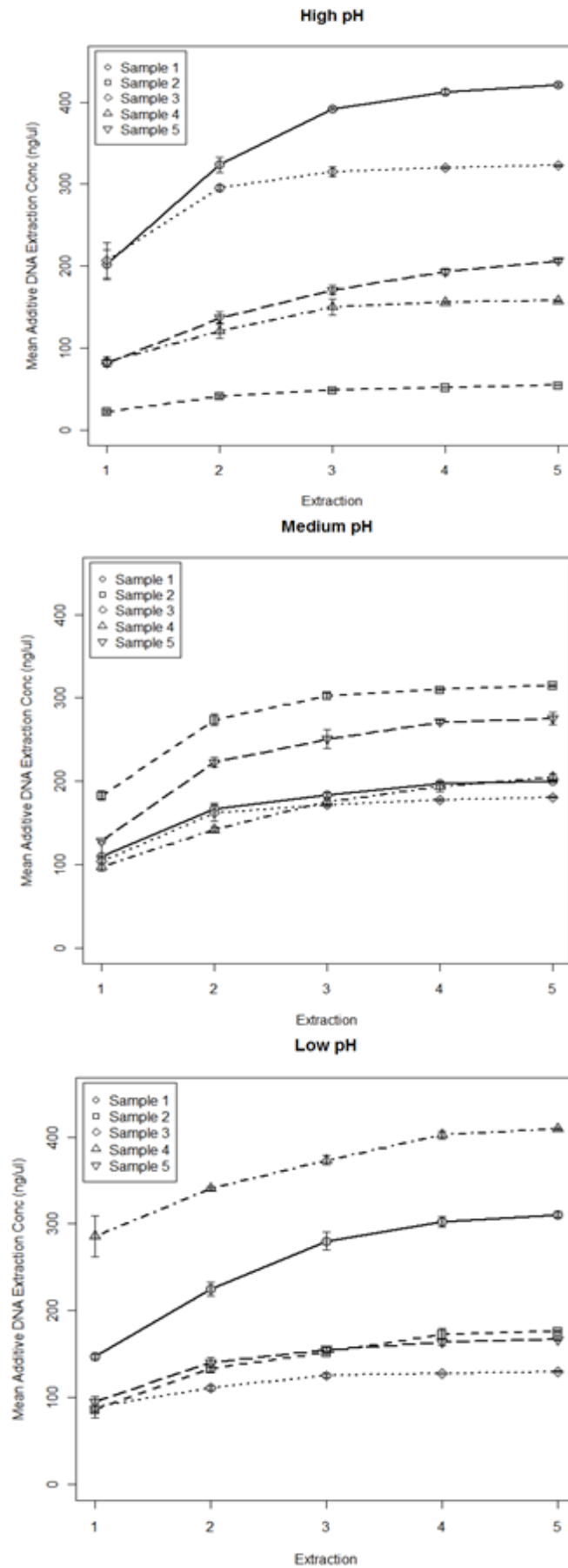


Figure 6: Additive soil nucleic acid concentrations following multiple extractions ($n = 5$) on replicated ($n = 3$) single aliquots of soils across a pH gradient (low pH $n = 5$, medium pH $n = 5$ and high pH $n = 5$). Error bars denote the standard error of the mean for each replicated sample extraction.

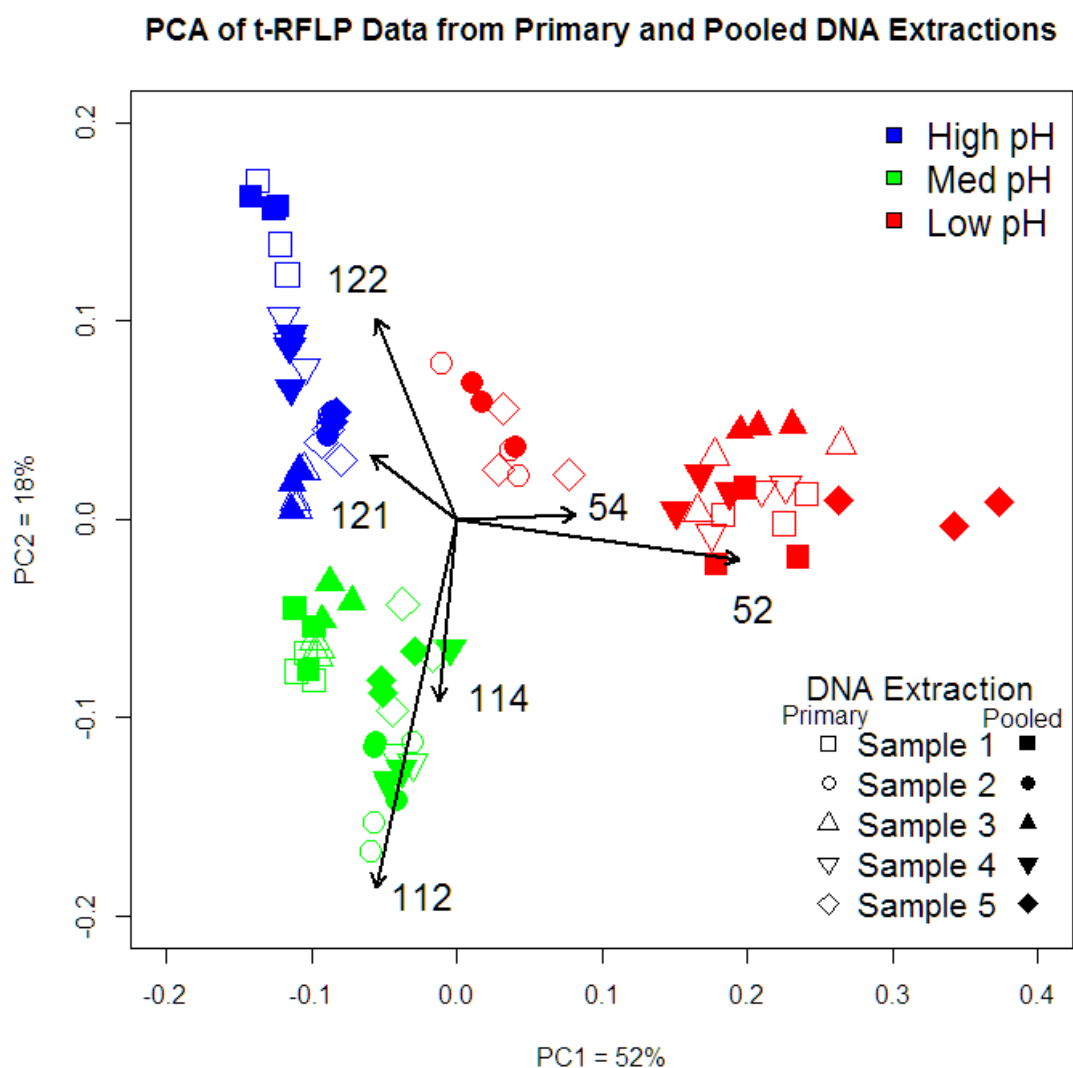


Figure 7: Principal component analysis of bacterial community structure from single extractions and 5 pooled repeated extractions (in equal volumes) from the same sample in triplicate replication. Vectors added to the plot highlight the dominant terminal restriction fragments (TRFs) predominantly responsible for the differences in bacterial community structure between the pH groups.

2.4 Discussion

2.4.1 Increased Nucleic Acid Yield from Clay Soils

Although the modified method was successful in increasing the yield of nucleic acids from clay soils, these soils maintained a tendency to produce low nucleic acid yields when compared with other soil types. This may be due to a combination of incomplete extraction, due to the ionic interactions between the polar DNA macromolecule and the charged surface of the clay particulates (Cai *et al.* 2006), and the physical entrapment of microbial cells within the clay substrate (Taylor *et al.* 2000). Further experimentation may help determine the effect of multiple extractions upon high clay content samples exclusively, as multiple physical and chemical lysis steps may release further nucleic acids entrapped within the substrate, or disrupt electrostatic interactions between nucleic acids and the clay particle surface. However, these factors are unlikely to produce a bias towards particular taxa of bacteria (i.e. DNA from all taxa will be affected equally). Thus it is likely that nucleic acid yields are representative of the microbial communities present within these high clay content soils. This study demonstrates that two minor modifications to the published method, the reduction of sample mass from 0.5 g to 0.25 g and an additional CTAB based freeze thaw step, significantly increase nucleic acid yield from clay soils with no significant effect upon nucleic acid purity. As clay soils typically produce nucleic acid extracts of such low quantity that they are unusable for downstream molecular analyses, a mean increase of $39 \text{ ng } \mu\text{l}^{-1}$ is a useful advance in methodology when compared with the original protocol. A reduction in sample mass may shift the equilibrium of dissociated DNA from the charged clay surfaces to the extraction buffer, and the preliminary freeze thaw step may disrupt the dense soil matrix reducing the entrapment of bacteria within. Furthermore the physical disruption of the soil matrix due to the freeze thaw step would increase the surface area exposed to subsequent physical and chemical lysis steps. Methods such as saturating the charged clay surface with eukaryotic DNA have been proposed (Summers, S., pers comm, 2010), but this prevents accurate nucleic acid quantification, and renders the sample potentially useless for any subsequent experiments focusing on eukaryotes. Other desorption methods have been suggested, including adding skimmed milk to compete with clay binding sites (Takada-Hoshino & Matsumoto, 2004). When this was combined with physical grinding of soil in liquid nitrogen, a large

increase in recovered nucleic acids was reported. However, including these extra steps increases the time taken to process samples, making the protocol impractical for use in large scale studies. Furthermore the inclusion of skimmed milk powder has led to increased contamination of samples by non-native eukaryotic DNA (Ikeda *et al.* 2008).

2.4.2 Factors Affecting Nucleic Acid Yield and Purity

In contrast to other published soil DNA extraction methods, no significant correlation was observed between soil carbon content and nucleic acid yield (Dequiedt *et al.* 2011; Zhou *et al.* 2006). This may have been due to the semi-quantitative nature of the method used to estimate nucleic acid concentrations in this study. In concurrence with Dequiedt *et al.* (2011), negative correlations between nucleic acid yield and soil pH and C:N ratio were observed but these correlations were significant but weak. Principal causes of the differences in nucleic acid yield across the soils studied may be the differences in below ground biomass between habitats, and the interactions between nucleic acid macromolecules and charged soil particles. Soil pH and carbon content have been shown to significantly affect the numerical abundance (Bååth *et al.* 2003; Mulder *et al.* 2005) and diversity (Fierer & Jackson, 2006) of below ground organisms and potentially explain variation in nucleic acid yield. However, unlike in Dequiedt *et al.* (2011), no calculation of soil microbial biomass was inferred directly from the extracted nucleic acid concentration as genetic material originating from plant and other members of the soil meso fauna would be included. Furthermore, as repeated nucleic extractions gave an increase in recovered nucleic acids, it can be hypothesised that quantifying nucleic acids is not an accurate proxy for determining soil biomass.

Soil texture has also been shown to significantly affect the numerical abundance of below ground organisms (Mulder *et al.* 2005), but only clay like soils produced significantly lower nucleic acid yields than all other soil types examined. This may be due to the continued adsorption of nucleic acids to charged clay particles as well as physical entrapment. Sandy soils have often been reported to contain low levels of microbial and plant biomass due to low levels of organic carbon input (Hassink, 1994) and low microbial carrying capacities (Veen *et al.* 1997). However within this study, large grain size soils

such as sands consistently produced high nucleic acid yields, possibly because of the lack of nucleic acid binding through charged electrostatic interactions. Finer textured soils such as loams and silty loams also produced high nucleic acid yields and it is known that soils such as these contain comparatively large numbers of below ground organisms as a result of protection from numerous adverse factors, including desiccation (due to higher water holding capacities), predation, gas diffusion, and toxic compounds (Ranjard *et al.* 2001; Dequiedt *et al.* 2011). Nutrient resources are also comparatively high within these soil types due to stabilization of organic matter (Wang *et al.* 2003), leading to a proliferation of microbial organisms in comparison with low nutrient soils.

Organic soils such as those supporting heath and bog communities, as well as moorland grass mosaics and upland wooded populations, produced high nucleic acid yields. In comparison, mineral soils such as those supporting crops and weed communities produced significantly lower nucleic acid yields. By definition, mineral soils contain higher levels of inorganic material than organic soil types, potentially resulting in lower levels of nucleic acid containing material per unit mass of soil. This is apparent in the IDW interpolation of nucleic acid yields across the UK, with the agricultural soils of the south east of England producing smaller nucleic acid yields than the organic soils found in parts of Wales and west Scotland. This hypothesis is further supported as other soil types containing high organic matter contents such as fertile and infertile grasslands also produced comparatively high nucleic yields, as was also seen by Dequiedt *et al.* (2011).

No large scale studies have examined the effect of climatic or edaphic variables upon extracted nucleic purity estimated by spectrophotometry. Organic soils are known to contain high levels of humic acids (Young *et al.* 2003) which reduce the 260nm /230 nm absorbance ratio. CTAB extraction procedures, as used here, have been shown to be one of the best methods for reducing contamination by these compounds (Zhou *et al.* 1996; Miller *et al.* 1999). Organic soils, such as peats and those supporting heath and bog, upland wooded, and moorland grass mosaic communities showed, reduced 260 nm/ 230 nm ratios in comparison with all other soil types. However, minimal dilution of the nucleic acid extract allowed consistent PCR amplification of extracted DNA (Chapter 3). Soil pH showed significant positive correlations with the 260

nm/ 230 nm ratios as low pH soils are generally organic in nature (i.e. peats) and high in humic acid concentration (Garner-Sillam *et al.* 1999). Positive correlations between C:N ratio and 260 nm/ 230 nm ratios were observed, suggesting that higher levels of contamination by substances such as humic acids occurred as the carbon content of soils increased. Again, higher carbon content soils are indicative of low pH organic soil types.

Compared with other European countries, the UK has a large coverage of soils with very high organic matter content for its size (Montanarella *et al.* 2006). Peat soils are globally important as carbon sinks (Davidson & Janssens, 2006) and as such, their below ground communities are of great interest to numerous branches of soil science. It is typically difficult to extract PCR amplifiable DNA suitable for molecular analysis from organic soils (Young *et al.* 2003). The method presented here is not only useful for obtaining consistent nucleic acid extracts from a wide variety of soil types, but also produces relatively clean nucleic acid extracts, suitable for downstream molecular analysis, from soils high in humic acid compounds. Interpolations of the 260 nm/ 230 nm ratios across the UK highlights areas which produced nucleic acid extracts with 260 nm/ 230 nm ratios below the desired 2.0. These are associated with areas known to have very high organic matter and humic acid contents (i.e. west Scotland and parts of Wales) (Montanarella *et al.* 2006).

The 260 nm/ 280 nm ratios were relatively constant across all soil types, floristic communities, and geographic space, and show minimal contamination by substances absorbing light at 280 nm such as phenolic compounds or proteins. Only weakly significant relationships between C:N ratio and 260 nm/ 280 nm ratios were observed, suggesting a minimal influence of soil nutrient status upon the contamination of nucleic acid extracts by compounds absorbing light at 280 nm. High carbon content soils are known to contain high levels of phenolic compounds (Freeman *et al.* 2001), and the slight positive correlation between pH and 260 nm/ 280 nm ratio can be explained by an increase in release of phenolic compounds such as *p*-hydroxybenzoic, vanillic, *p*-coumaric, ferulic and syringic acids, *p*-hydroxybenzaldehyde and vanillin from soils with increasing soil pH (Whitehead *et al.* 1981). The consistent 260 nm/ 280 nm absorbance ratios close to the ideal value of 1.8 indicate that carryover of phenol during the extraction procedure should be minimal. Furthermore

differences in results due to experimenter bias were minimised as extractions were carried out by a single researcher.

2.4.3 Effects of Multiple Extractions Across a pH Gradient

In concurrence with Feinstein *et al.* (2009), an increase in the nucleic acid yield was observed over multiple extractions performed upon the same soil sample. Approximately 50 % - 70 % of the genetic material obtained over 5 sequential extractions was obtained within the first extraction, and the first 3 extractions capture the majority of the total nucleic acids within a sample. In contrast to the findings of Feinstein *et al.* (2009), there is little evidence that the commonly reported pH effect upon soil bacterial community dissimilarity and diversity is a result of incomplete nucleic acid extraction. In fact here the pH effect is, in a single example, increased with multiple extractions and thus we propose that only a single extraction is needed when community profiling methods with a low taxonomic resolution (e.g. t-RFLP) are used. To hypothesise, as a large proportion of the available nucleic acid is obtained during the first extraction, any subtle differences in actual taxa abundance resulting from subsequent extractions are generally below the detection limit of community profiling methods such as t-RFLP. However, the use of next generation pyrosequencing to provide detailed taxonomic insights into soil microbial populations is rapidly gaining popularity (Chistoserdova, 2010). Therefore, as slight changes in community structure were detected as a result of multiple extractions in a small number of samples, it would be prudent to suggest using a minimum of 3 sequential pooled extractions from the same soil sample as a basis for studies utilising this technology. An increase in the proportional abundance of actinobacterial taxa over multiple extractions observed in the study by Feinstein *et al.* (2009) could be due to the Gram positive nature of this lineage. Chemical lysis steps used in commercial high throughput soil DNA extraction kits may fail to fully lyse cells with high peptidoglycan contents and thus result in a disproportionate release of actinobacterial genetic material across multiple extractions. The use of a physical lysis step (i.e. bead beating) should overcome this problem as it has been shown to successfully extract high levels of Gram positive bacterial DNA in gut microbiome studies (Pryde *et al.* 1999; Zoetendal *et al.* 2001) and is thought to result in effective lysis of all soil organisms (Yeates *et al.* 1998).

2.4.4 Conclusions

In conclusion, reported here is a relatively quick, cost effective, and simple nucleic acid extraction method that results in nucleic acid yields from a wide range of soils with varying physiochemical properties suitable for use with high throughput studies. Furthermore this method shows potential for use with studies examining key soil microorganisms of diverse phylogenies. Minimal biases were detected as a result of incomplete DNA extraction when extracts were examined by t-RFLP. Also, 3 multiple extractions capture a larger proportion of nucleic acids present in a sample than a single extraction, and thus the use of three extractions should be considered for studies based around high throughput next generation sequencing.

2.4.5 Further work

Further studies have shown that this method is suitable for use with other organisms such as the Archaea, fungi (Chapters 5 - 6), and other members of the soil meso and macro fauna (unpublished metagenomic and next generation sequencing assessments). Further experimentation may be required to assess the effect of multiple extractions upon the efficiency of nucleic acid extraction from other members of the soil biota, but for soil fungi, other studies have not seen significant biases introduced through incomplete extraction (Feinstein *et al.* 2009).

Studies of the soil transcriptome require efficient RNA extraction. Although the effect of the additional freeze thaw step and reduction of sample mass upon RNA extraction efficiency was not directly assessed, the published version of this method reported rapid co-extraction of both DNA and RNA from soil samples. There is little evidence to suggest that the modification to this method would have a large detrimental affect upon RNA yields, and nucleic acids of lengths characteristic of RNA were detected through gel electrophoresis (data not shown). Further work may include examining the extraction efficiency of different RNA types and their stability after extraction in comparison to other soil nucleic acid extraction methods, as well as the effect of climatic and edaphic variability upon RNA yield and purity.

Chapter 3: Landscape Scale Terminal Restriction Fragment Length Polymorphism Analysis of Soil Bacterial Communities across the UK

3.1 Introduction

3.1.1 Background

Bacteria are ubiquitous in the soil environment, and form some of the most diverse ecological systems on the planet (Fierer & Jackson, 2006). Estimates of bacterial species numbers within soil samples of varying quantities can be calculated using experimentally derived genetic diversity measures (such as from t-RFLP, DGGE, and 16S rRNA sequencing methods) combined with statistical modelling of taxa abundance curves, and known quantities of the most abundant and least abundant taxa detectable (as in Curtis *et al.* 2002). Species estimates per gram of soil are generally regarded to be within the range of $10^3 - 10^7$ (Curtis *et al.* 2002; Gans *et al.* 2005; Fierer *et al.* 2007; Torsvik *et al.* 1990). Due to the large functional potential associated with this diversity (Nannipieri *et al.* 2003), soil bacteria are regarded as key organisms in practically every biogeochemical cycle (Nacke *et al.* 2011). As a result, the continued functioning of terrestrial ecosystems is reliant upon soil bacteria. However, a basic understanding of how populations of these organisms are structured is lacking due to issues associated with their enormous diversity (Torsvik *et al.* 1990; Fierer & Jackson, 2006), cultivability (Janssen *et al.* 2002), and methodological limitations to molecular based assessments (Zhang & Xu, 2008).

Until recently the term “biogeography” was predominantly applied to larger organisms, such as plants and animals, to describe differences in species abundances over an area in response to a combination of biotic and abiotic factors (Fierer, 2008). It has been argued that biogeography may not apply to organisms smaller than 1 mm in diameter, due to an assumed ubiquitous and homogeneous dispersal potential (Finlay, 2002). However, if there were no biogeographic effect for soil microbial populations, then all soil microbes would

be homogeneous at all spatial scales (Fierer, 2008). This has repeatedly been shown not to be the case (Lauber *et al.* 2009; Acosta-Martinez *et al.* 2008; Rousk *et al.* 2010). Clearly then, an understanding of the drivers which shape soil bacterial populations is an essential step in understanding how ecosystems operate (Martiny, 2006).

Soil bacterial community structure, composition, and biomass are known to vary both locally and globally in response to a number of biotic and abiotic factors, for example, soil nutrient status (Tiquia *et al.* 2002; Chu *et al.* 2007), plant communities (Marschner *et al.* 2001; Westover *et al.* 1997; Kourtev *et al.* 2002), soil pH (Fierer & Jackson 2006), soil texture (Girvan *et al.* 2003; Johnson *et al.* 2003), soil depth (Griffiths *et al.* 2003; Fierer *et al.* 2003), oxygen gradients (Lüdemann *et al.* 2000), heat gradients (Norris *et al.* 2002), land use (Lauber *et al.* 2008), pollution levels (Smit *et al.* 1997; Müller *et al.* 2001), soil matrix disruption (Lupwayi *et al.* 1998), seasonal influences (Smit *et al.* 2001), drying and re-wetting stresses (in conjunction with plant cover) (Fierer, & Schimel., 2002), and protozoan grazing (Rønn *et al.* 2002). In fact, most studies examining bacterial community structure, composition, diversity, or biomass within soils exhibiting different properties have reported significant shifts in ecological descriptors of bacterial populations such as measures of taxa richness and evenness. As most edaphic properties are highly co-correlated, for example changes in land usage will affect the physicochemical properties of a soil, it is unclear which variables are the most important determinants of bacterial diversity, community structure, and composition. Many of the aforementioned studies have identified key drivers of soil bacterial community composition, but are focused at the local scale rather than the landscape scale.

More recently, high-throughput molecular based studies of soil bacterial community structure (Fierer & Jackson, 2006; Lauber *et al.* 2009; Dequiedt *et al.* 2011) have been performed, providing an improved comparative investigation of the factors which shape soil bacterial communities at larger spatial scales. Pioneering landscape scale surveys of bacterial biogeography have been limited in terms of spatial resolution (e.g. that of Fierer & Jackson 2006, where 98 soil samples were used to represent the entire Americas), while others have higher spatial resolutions, but examine bacterial biogeography over land masses with a unique composition of soil types (Dequiedt *et al.* 2011).

Furthermore, these studies are in disagreement over the principal driver of bacterial diversity, the former suggesting soil pH, the latter inferring soil type and land cover.

Terminal restriction fragment length polymorphism (t-RFLP) is a popular molecular community profiling method used to discriminate between microbial taxa based upon sequence variability of a particular taxonomic marker gene. Using t-RFLP, a number of studies have shown soil pH to have the greatest effect on soil bacterial diversity over all other environmental or spatial factors (Nicol *et al.* 2008; Fierer & Jackson, 2006). Fierer & Jackson (2006) showed that, in 98 soils sampled from across the Americas, bacterial diversity exhibited a unimodal response to soil pH, with the highest bacterial diversity found in soils around pH 6-7. Other studies that have focused upon the effect of pH on bacterial community structure over individual habitat types, have reported different effects of pH upon diversity (Chu *et al.* 2010). This highlights the need for further in-depth, large scale surveys of bacterial communities to establish fundamental ecological patterns.

The UK has the most varied geology of any land mass for its size (Bennison & Wright, 1969) and thus the parent material of soil is equally as varied. Large environmental gradients can be found within the UK from the colder harsher climates of north west Scotland, to the comparably drier and warmer areas of the south and south west UK. This, in turn, results in a wide variety of habitats and soil conditions across this landmass. Therefore the UK provides the perfect model system to highlight the relationships between numerous biotic and abiotic conditions and soil bacterial diversity and community structure in an effort to highlight landscape scale patterns in soil bacterial biogeography.

3.1.2 Aims

The aims of this work are to elucidate the main environmental drivers of bacterial community structure and diversity across the UK using a 16S rRNA gene t-RFLP approach. A second aim is to construct a basic predictive model using a recursive partitioning method to define the species environment relationship. A final aim is to generate interpolated maps representing the main changes in bacterial community structure across the UK.

3.2 Methods

3.2.1 Sample Selection and DNA Extraction

Soils were collected as part of the Countryside Survey 2007. For further information see sections 2.3.1 (Figure 2) and 2.3.2.

3.2.2 Polymerase Chain Reaction

PCR and T-RFLP reactions were set up as described in section 2.2.5.

3.2.4 Statistical Analysis

All statistical analyses were performed in the programming environment R (Ihaka & Gentleman, 1996) using packages available under GNU licence from the CRAN repository. These packages included vegan (Oksanen *et al.* 2011), MASS (Venables & Ripley, 2002), BiodiversityR, (Kindt, 2005), and cluster (Maechler, 2011). Prior to statistical analysis involving numerical environmental variables, all were scaled to have a mean of 0 and a standard deviation (SD) of 1 for correct unit comparison. For instances where mean values are plotted, the standard error of the mean is displayed as error bars where appropriate. Dummy variables were used to encode categorical variables where a numerical variable was essential.

The richness and evenness of bacterial communities (alpha diversity) at each site were calculated using Simpson's index of diversity ($1 - D$) as defined by equation 1, where p_i represents the proportional abundance of the i th TRF within a sample.

Equation 1: Simpson's Diversity index (alpha diversity)

$$\alpha = 1 - \sum_{i=1}^S (p_i)^2$$

Beta diversity is reported as the average variability in diversity between particular groups of samples. Beta diversity was calculated using equation 2, where p_{ij} is the proportional abundance of the i th TRF within a sample, and \bar{p}_i is the mean abundance of the i th TRF across all samples within the group.

Equation 2: Mean variance in Diversity (beta diversity)

$$\beta = \sum (p_{ij} - \bar{p}_i)^2$$

Gamma diversity, or total diversity, was reported as the sum of alpha and beta diversity.

Equation 3: Total diversity (gamma diversity)

$$\gamma = \alpha + \beta$$

Non-metric multidimensional scaling (NMDS) was used to analyse bacterial community dissimilarity. It is generally regarded as one of the most robust unconstrained ordination methods in community ecology (Minchin, 1987). The metaMDS function within the vegan package in R was used to create NMDS ordinations based on TRF proportional abundance data. A number of iterations with random starts were performed (n = 20) to avoid the calculation becoming stuck in a local optimum, as can frequently occur with the NMDS method. The envfit function in MASS was used to fit environmental variables to the plot using a least squares regression method over 1000 permutations as a test for significance. Isoclines depicting predicted values of a single environmental variable were added using a general additive model (GAM) approach. Where this is reported, the R² value indicates the goodness of fit of the isoclines to the data points. A stress plot was created to analyse the correlation between calculated community dissimilarity between sites and plotted distance on the ordination (Appendix F).

Where the term “DCA axis” is used as an environmental variable, this refers to particular axis scores of a de-trended canonical correspondence analysis representing the total above ground plant communities present at each sample site. The DCA was calculated by the Countryside Survey 2007 statistical analysis team.

Recursive multivariate partitioning analysis was performed using the mvpart function in R to describe and predict relationships between multispecies data and environmental characteristics (De'ath, 2002). Multivariate regression trees were calculated to 1 standard error, and a cross validation method was

employed using one half of the data to create the model and the other to test it. The multivariate regression tree was created using the Cluster package (Maechler *et al.* 2011) in R to describe the taxa-environment relationship for soil bacterial communities. This was done in an attempt to create a broad scale predictive model characterising typical bacterial t-RFLP profiles generated for soil types with known physiochemical properties. The full t-RFLP dataset was repeatedly split based upon rules governed by all environmental variables to minimize community dissimilarity within each sub group. A mean t-RFLP profile was displayed at each terminal node.

Inverse distance weighted interpolation was used to predict NMDS axis one scores of bacterial communities for un-sampled locations across the UK. This was achieved by using information from data points surrounding that location to a greater or lesser degree as dictated by distance from the un-sampled location. IDW interpolation was carried out using the default settings in ARCGIS.

To identify specific t-RFLP peaks of interest, “*in-silico*” digestion of a global 16S rRNA database was performed. Arb (Ludwig *et al.* 2004) was used to create a database of 33168 high quality, chimera checked, near full length 16S rRNA sequences from the greengenes (DeSantis *et al.* 2006) online sequence repository. All sequences were screened to ensure a soil based isolation source prior to “*in-silico*” digestion using the Arb integrated program TRFCut (Ricke *et al.* 2005). The targeted region of each sequence was identified using a global primer search of all aligned sequences, and sequences exhibiting large gaps within the cut site region were discounted from further analysis. When identifying peaks, size differences of a few base pairs were tolerated to account for small discrepancies in experimental procedures from lab to lab.

Where environmental variability was calculated with increasing soil pH, all environmental data was partitioned into 6 discrete classes representing pH values ranging from 3 to 9 in single pH increments. Euclidean distance matrices were then created for each pH class using the vegdist function in the R package vegan (Oksanen *et al.* 2011), the mean of which was plotted for each pH class along with the standard deviation of each matrix as error bars.

3.3 Results

Of the 1114 soil samples that underwent molecular analysis, 104 were removed from subsequent analysis due to poor quality t-RFLP runs ($n = 38$), as highlighted by kmeans clustering, and/or incomplete environmental data ($n = 68$).

The kmeans clustering algorithm was used again to partition the t-RFLP profiles into two groups based upon dissimilarity of bacterial communities. A matrix plot was generated to identify key broad scale changes in proportional abundance of TRFs between the two groups (Figure 8). Rows represent individual samples, and columns represent the terminal restriction fragments, with their proportional abundance indicated by band intensity. Figure 8 highlights that specific terminal restriction fragments can be found at higher proportional abundances within a particular kmeans group when compared with the other kmeans group. In particular, fragments 52, 110, and 226-229 bp in length are found at higher proportional abundances in kmeans group 1, whereas 112 and 121 bp in length are predominantly found in kmeans group 2.

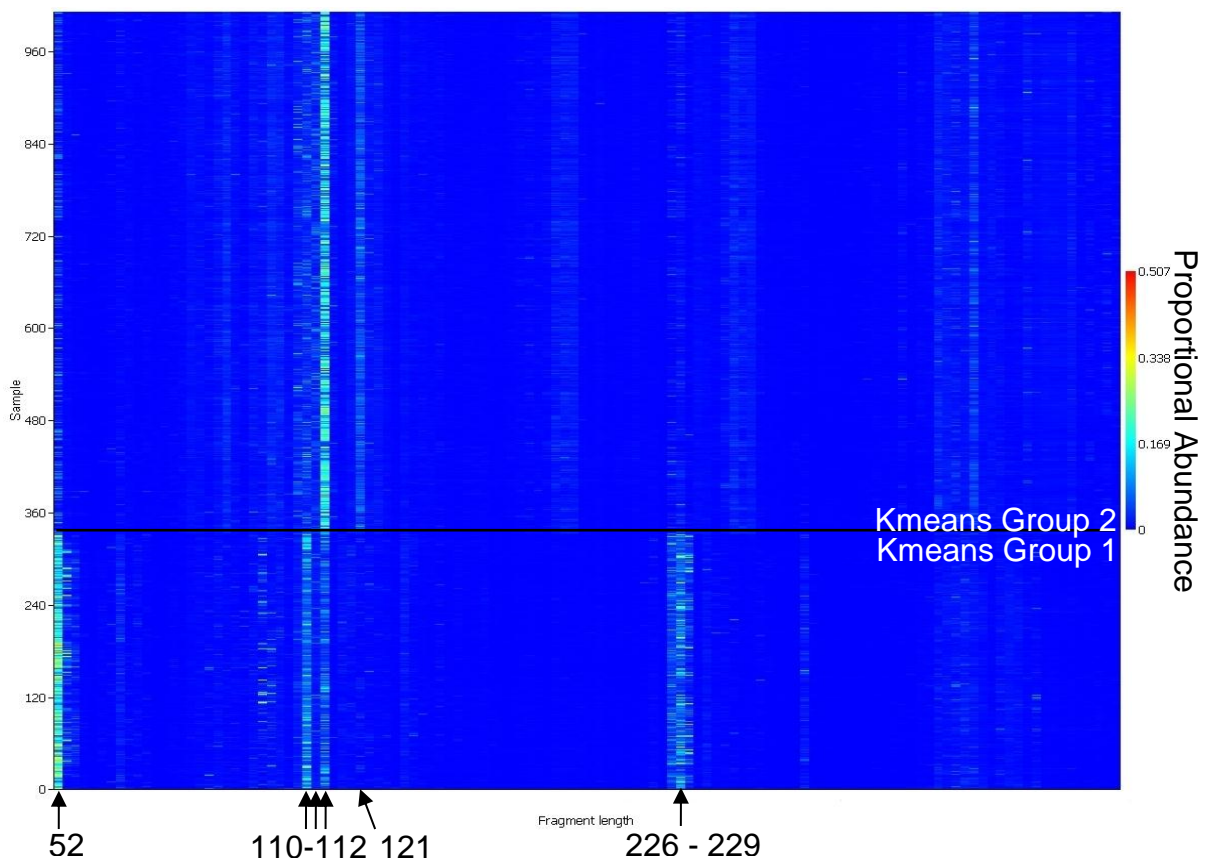


Figure 8: Matrix plot showing the proportional abundance of TRFs ordered into kmeans group 1 (bottom) and 2 (top). TRFs identified as dominant within either group are highlighted by arrows. The TRFs 52, 110 and 226-229 nt in length were identified as the dominant TRFs within kmeans group 1. Dominant TRFs within Kmeans group 2 were 112, and 121 nt in length.

A preliminary principal component analysis confirmed the existence of two distinct communities, with kmeans group 2 being dominated by TRFs 112 nt and 121 nt in length. Kmeans group 1 was characterised by high abundances of TRFs 52 nt long (Figure 9). These terminal restriction fragments were shown to be important determinants of each kmeans group through the addition of vectors (in the form of arrows) to the PCA plot. The vectors represent the magnitude and direction of the effect each terminal restriction fragment has upon the separation of the data points across the principal components displayed. TRFs 52, 112 and 121 bp in length had the longest vectors and were thus deemed the most important determinants of the two kmeans groups. These rudimentary analyses highlighted a biogeographic effect as bacterial communities were found to fall into two distinct groupings depending on differences in the abundances of different TRFs.

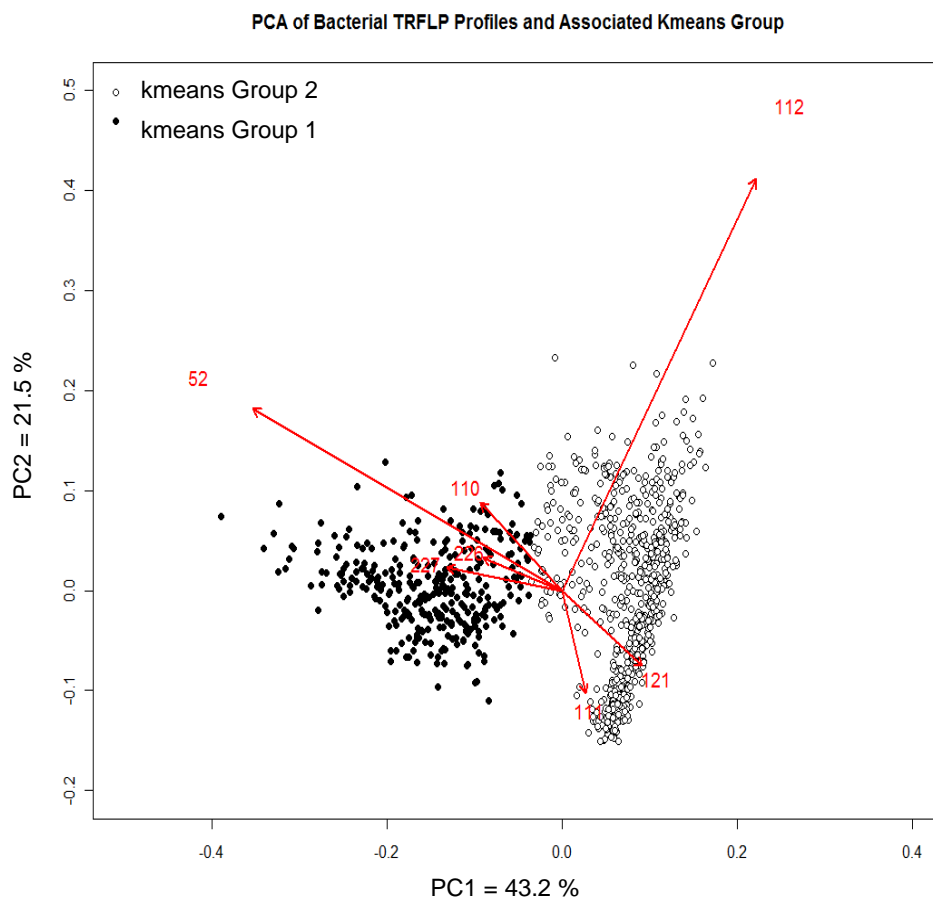


Figure 9: Principal component analysis showing the TRFs responsible for separating the Kmeans group 1 and 2 bacterial communities. The seven TRFs that had the greatest effect upon the separation of data points are shown. TRFs 52, 110, 227 and 226 were typical of kmeans group 1, while TRFs 111, 121 and 112 were typical of kmeans group 2. TRFs 52, 112 and 121 were highlighted as the most important TRFs causing the separation of the data points due to the length and direction of vectors. PC axis percentage denotes the percentage of variance explained by each axis.

To identify any environmental variables responsible for this biogeographic effect, NMDS analysis of bacterial t-RFLP data was carried out (Figure 10). Table 3 shows the relative effect of environmental variables upon bacterial community dissimilarity within the plot, reported as the R^2 value from a least squares regression upon the first and second axis scores of the NMDS ordination.

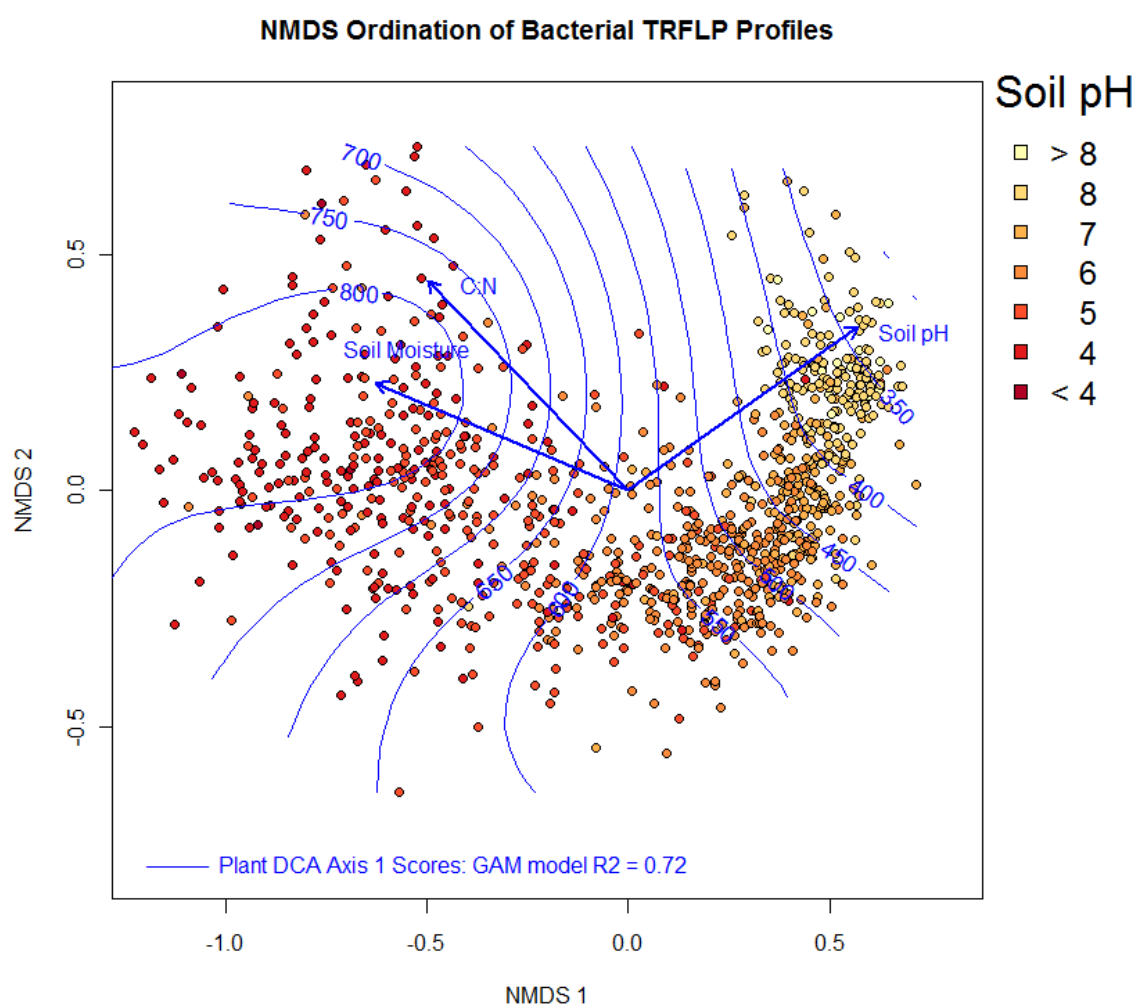


Figure 10: Non-metric multidimensional scaling (NMDS) plot showing the overall bacterial community dissimilarity re-calculated from Griffiths *et al.* (2011) with additional environmental variables. Point colour represents soil pH of each sample, and the magnitude and direction of arrows represent the effect of environmental variables upon the separation of the data points. Isoclines represent predicted values of plant DCA axis 1 scores and show the association between plant community dissimilarity and bacterial community dissimilarity. Appendix F shows the stress plot for this ordination.

Environmental Variable	R ²	p value
Soil pH	0.7788	0.001
Dominant Plant Species	0.6566	0.001
Plant DCA Axis 1 Scores	0.6167	0.001
A V C	0.6145	0.001
Broad Priority Habitat	0.6102	0.001
Soil Moisture content (%)	0.5334	0.001
Carbon:Nitrogen	0.533	0.001
Loss on Ignition (g)	0.511	0.001
Carbon Content (%)	0.4995	0.001
Nitrogen Content (%)	0.41	0.001
Soil Texture	0.3824	0.001
Annual Rainfall (mm)	0.3461	0.001
Annual Temperature (°C)	0.3362	0.001
Annual Sunshine (Hrs)	0.3274	0.001
Annual Cloud Cover (Hrs)	0.3051	0.001
Altitude (m)	0.2646	0.001
Plant DCA Axis 2 Scores	0.2469	0.001
Phosphorus Content (5)	0.0466	0.001
Plant DCA Axis 3 Scores	0.0015	0.533

Table 3: Results of least squares regression of scaled environmental variables upon NMDS ordination in Figure 10 ordered by decreasing R² value. Soil pH explained the largest proportion of the bacterial community dissimilarity, followed by plant community and habitat indicators. All other environmental variables were found to be significant but to a lesser degree (except DCA axis 3 score).

Soil pH had the strongest correlation of all environmental variables with bacterial community dissimilarity at the landscape scale, accounting for 77 % of the variance within the plot. However, all other environmental variables, apart from plant DCA axis 3 scores, were significantly correlated with bacterial community structure. Points on the ordination were coloured to represent the pH of the sample (rounded to the nearest integer) as this was the main effect causing separation of the data points across the first NMDS axis. Isoclines representing predicted values for the plant DCA axis 1 scores were added to the plot in order to examine the linearity of this effect. The effects of other environmental variables were added to the plot as vectors, where the position and direction of the arrows signifies the relative magnitude of the effect (Figure 10).

Communities represented by points towards the left of the ordination, with the most negative first axis NMDS scores, typically came from low pH environments with high moisture contents and high carbon to nitrogen ratios. DCA axis 1 scores suggest these samples had distinct floristic communities (high plant DCA

axis 1 scores). In contrast, samples to the right of the ordination typically came from areas with a comparatively high pH, lower moisture contents, and lower carbon to nitrogen ratios. Again plant DCA axis 1 scores decreased when moving from left to right on the ordination suggesting a shift in plant communities with pH across the first axis.

Bacterial community structure data was mapped to display landscape-scale changes in bacterial diversity. This was achieved by inverse distance interpolation of the first axis NMDS scores across a map of the UK, using ARCGIS and the geospatial analyst plugin (Figure 11a). To highlight changes in bacterial community structure with respect to plant community dissimilarity, an IDW interpolation of the first axis scores of the plant DCA was performed (Figure 11b). North west areas of Scotland were dominated by communities from the left of the NMDS ordination (negative first axis NMDS scores), with low soil pH values and distinct plant communities. These soils are known to be predominantly from upland habitats and contain high amounts of organic matter. Similar bacterial communities were found in areas of Wales from similar habitats. Moving towards the typically agricultural, high pH soils of the south east of England, a change in bacterial community structure was observed. Again, the plant communities found in the south east of England were distinctly different from those in the north of the UK, being characterised by agricultural communities.

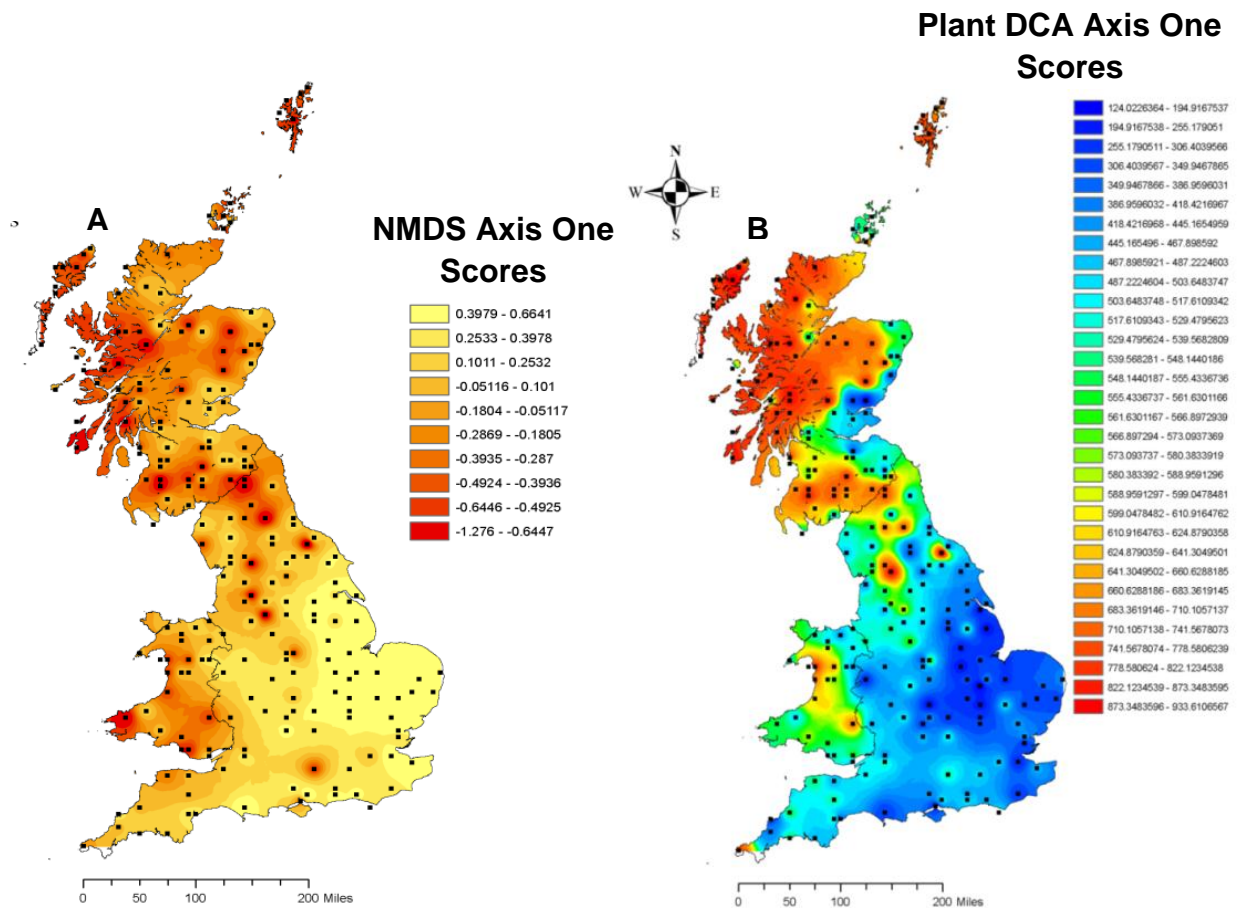


Figure 11: a) Inverse distance weighted (IDW) interpolation of NMDS first axis scores across the UK. Bacterial communities from the left of the ordination in Figure 10 were found predominantly in areas known to have acidic soil types with high organic contents. Those situated to the right of the ordination were seen in habitats such as the agricultural land of south east England. b) IDW interpolation of the plant DCA axis one scores across the UK. Areas such as the south east of England were shown to have distinct floristic communities to soils found in the north west of Scotland.

Focusing on the environmental gradients found when moving from the west of Scotland to the east, the higher pH environments and different plant communities found on the east coast also had a distinct association with bacterial community structure over this comparatively small geographic space. These communities were shown to be more similar to the communities found in the higher pH soils of the south east of England than to those of the upland areas of the west coast of Scotland.

3.3.1 Identification of t-RFLP Peaks using *in-silico* Digestion of a Global 16S rRNA Gene Database

Figure 9 highlighted the three most important TRFs with respect to producing a community shift between the Kmeans groups. In order to give taxonomic identifications to these TRFs, *in silico* digestion of a large global 16S rRNA

database was performed. Sequences that produced *in silico* generated t-RFLP fragments around 52 bp in length were predominantly identified as Acidobacteria (71 %) (Figure 12a), while those producing fragments of around 112 bp were predominantly Alphaproteobacteria (71 %) (Figure 12b). Sequences that were cut at approximately 121 bp were all actinobacterial taxa (83 %) (Figure 12c).

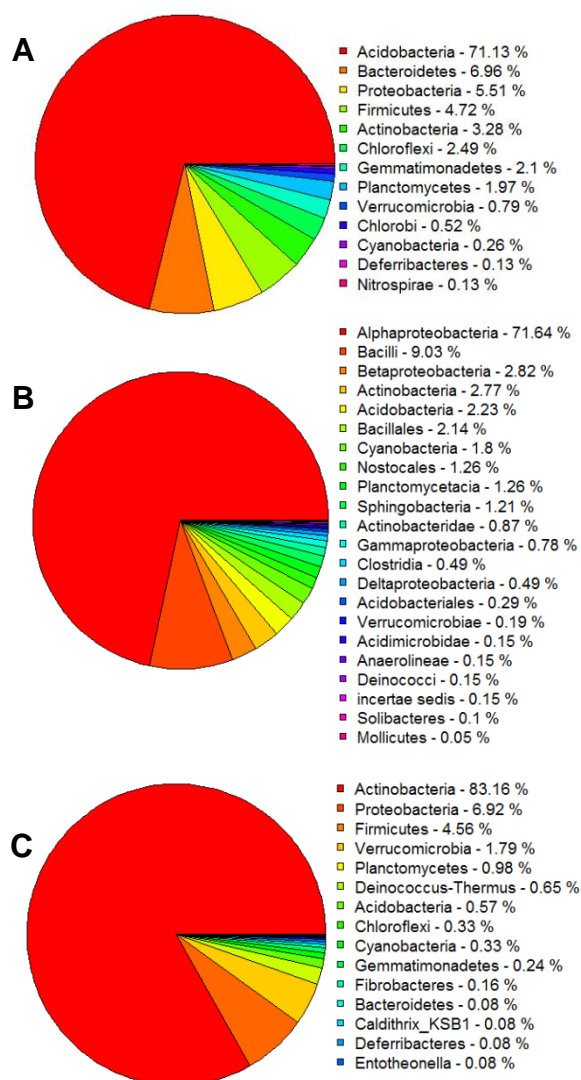


Figure 12: *In-silico* digestion of high quality full length 16S rRNA gene sequences from soils (n = 33168) to identify t-RFLP peaks characterizing UK soil bacterial communities identified in figure 10. A) TRFs 51 – 59, B) TRFs 111 – 116, C) TRFs 121 – 125.

3.3.2 Effects of Soil pH upon Bacterial Diversity

As soil pH was determined to be the principal driver of bacterial community structure and diversity above all other environmental variables (Appendix G), the effect of soil pH upon soil bacterial diversity was examined in detail.

Samples were binned into 6 discrete pH classes (pH 3.5 to 8.5) and alpha diversity was examined by calculating the mean Simpson's diversity ($1 - D$) for each pH bin (Figure 13).

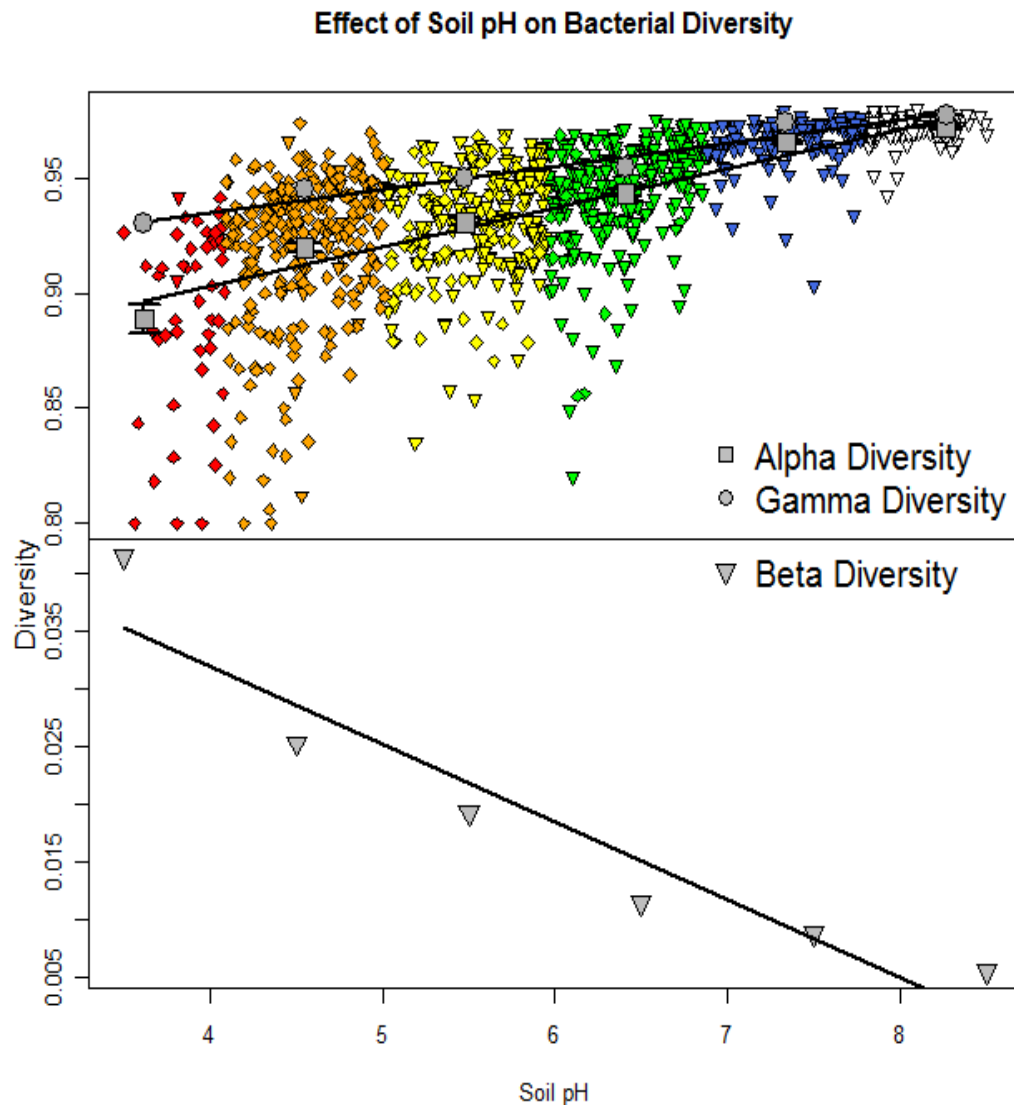


Figure 13: Relationships between soil bacterial diversity and soil pH. Soil pH was categorised into discrete bins ($n = 6$). In the top graph diamonds represent organic habitats: Heath and bog; Moorland Grass Mosaics; and Upland wooded communities. Inverted triangles represent improved habitats: Crops and Weeds; Fertile Grassland; Tall Grass and Herbs; Infertile Grassland; and Lowland Wooded. In the bottom graph inverted triangles represent beta diversity scores for each pH bin. Redrawn from Griffiths *et al.* (2011).

A significant increase in diversity was observed with increasing pH ($R^2 = 0.42$, $p < 0.001$). However, unlike the bacterial biogeography of the Americas (Fierer, *et al.*, 2006), a unimodal relationship was not seen. The mean variance in diversity was calculated to investigate beta diversity. Variation in bacterial community structure between soils was found to be greater at low pH, and significantly decreased with increasing pH. Despite the negative correlation between soil pH

and beta diversity, overall gamma diversity was still found to significantly increase with increasing soil pH.

The decrease in beta diversity with increasing pH suggests a higher degree of variability among bacterial communities in low pH environments than among those in high pH environments. A potential reason for this is the level of environmental variability inherent to each pH class. To investigate this issue, the environmental variability within each soil pH bin was calculated. All environmental variables were binned into 6 discrete groups based upon pH (3 – 9 in single pH unit increments) and a Euclidean distance matrix was calculated for environmental variables within each pH bin (Figure 14).

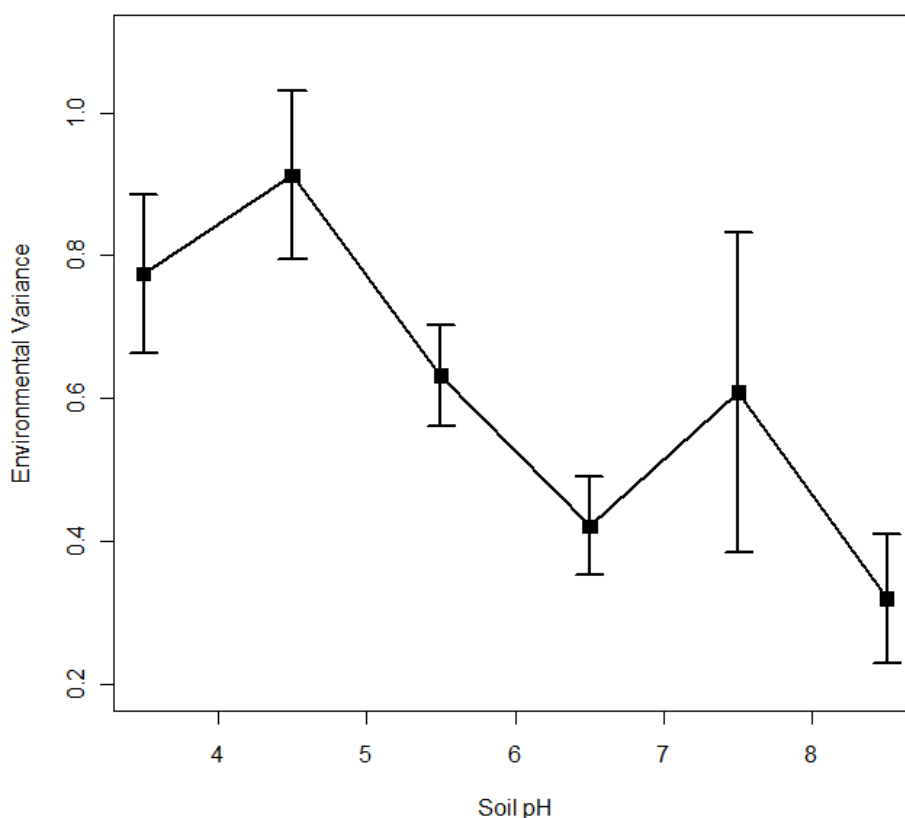


Figure 14: Environmental variability decreases as soil pH increases. All numerical environmental conditions were scaled to have a mean of zero and a standard deviation of 1 for unit comparisons. The variability was calculated for each environmental variable within a pH class and displayed as a mean value. Error bars describe the standard error.

Similarly to beta diversity, a general negative trend was observed when environmental variability was correlated with soil pH. This suggests a larger number of potential niches available for exploitation by bacterial communities within low pH habitats. This provides evidence to explain why the inter

community variability (beta diversity) decreases with increasing pH, but alpha diversity exhibits the opposite trend.

To investigate the distribution of particular groups of soil bacteria across the UK, the relative abundances of t-RFLP peaks 52, 112 and 121 were examined in relation to soil pH (Figure 15a-c). T-RFLP fragments 52 nt in length (identified as acidobacterial taxa) were at higher abundances in low pH environments and decreased with increasing pH. A unimodal response to pH was observed in the proportional abundance of TRFs of 112 nt in length (Alphaproteobacteria), with maximum proportional abundances found in soils with approximately neutral pH levels. The proportional abundance of TRFs 121 nt in length (Actinobacteria) increased linearly with increasing soil pH; however, the maximum proportional abundance of these TRFs was smaller than that of TRFs 52 nt and 112 nt in length.

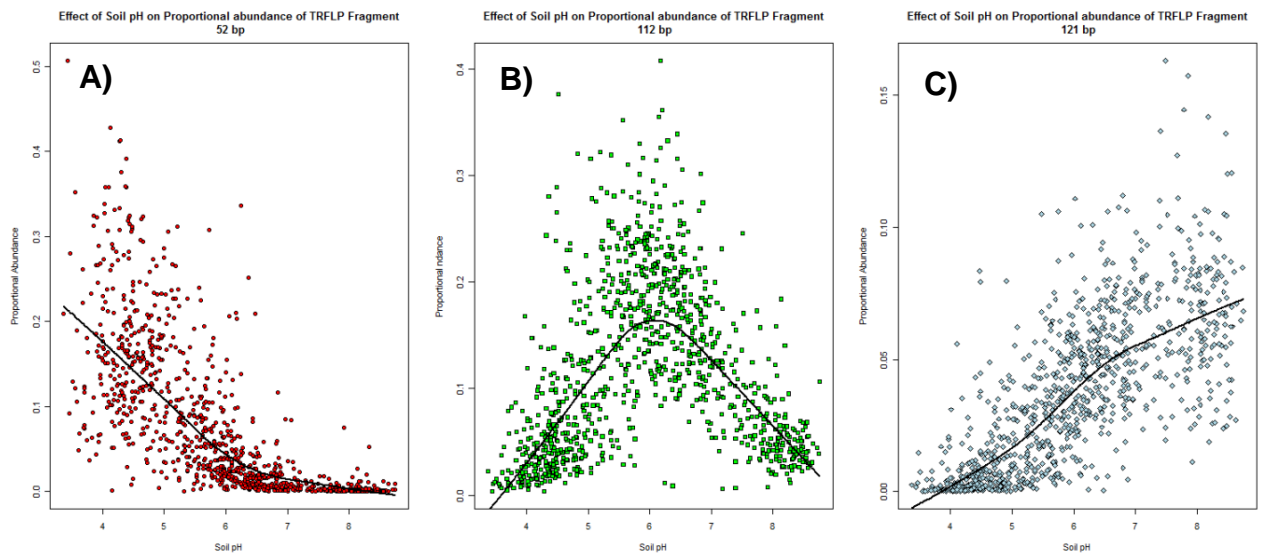


Figure 15: The proportional abundance of specific TRFs at increasing pH levels. A) Proportional abundance of TRF 52 bp in length decreases with increasing pH, B) TRF 121 bp in length shows a unimodal response to increasing pH. C) TRF 121 increases with increasing pH with maximum levels found within the soils with the highest pH values. Lines were fitted to show the general trend of data using a locally weighted polynomial regression fit.

In an effort to understand the changes in proportional abundance of specific important TRFs, IDW interpolations were undertaken for TRFs 52, 112, and 121 across the UK as they were highlighted as important taxa responsible for bacterial community structure differences in Figure 9 (Figure 16). TRF 52 was most abundant in the low pH upland habitats found on the west coast of Scotland and in areas with similar biogeochemical properties in south Wales. A

single area in the south east of England was shown to have higher abundances of TRF 52 bp in length than the surrounding areas. This was also highlighted in Figure 12a by a dark red area signifying bacterial communities to the left of the NMDS ordination (Figure 10). This is likely to be because the high organic and low pH soil found within the New Forest has more similar biogeochemical properties to soils of the upland habitats found in the Scottish west coast than to those of the surrounding area. TRF 52 is evidently characteristic of soils with these properties and largely responsible for the pH effect observed in the NMDS ordination. TRF 112 was generally found in high abundances in soils with a more neutral pH towards the middle of the country and in areas with typically “improved” soil types, while TRF 121 was typically found in high abundance in alkaline soils having high calcium content and in improved soil types (e.g. the agricultural land of the South Downs). Furthermore this TRF was seen in very low abundances in acidic areas such as the north west of Scotland.

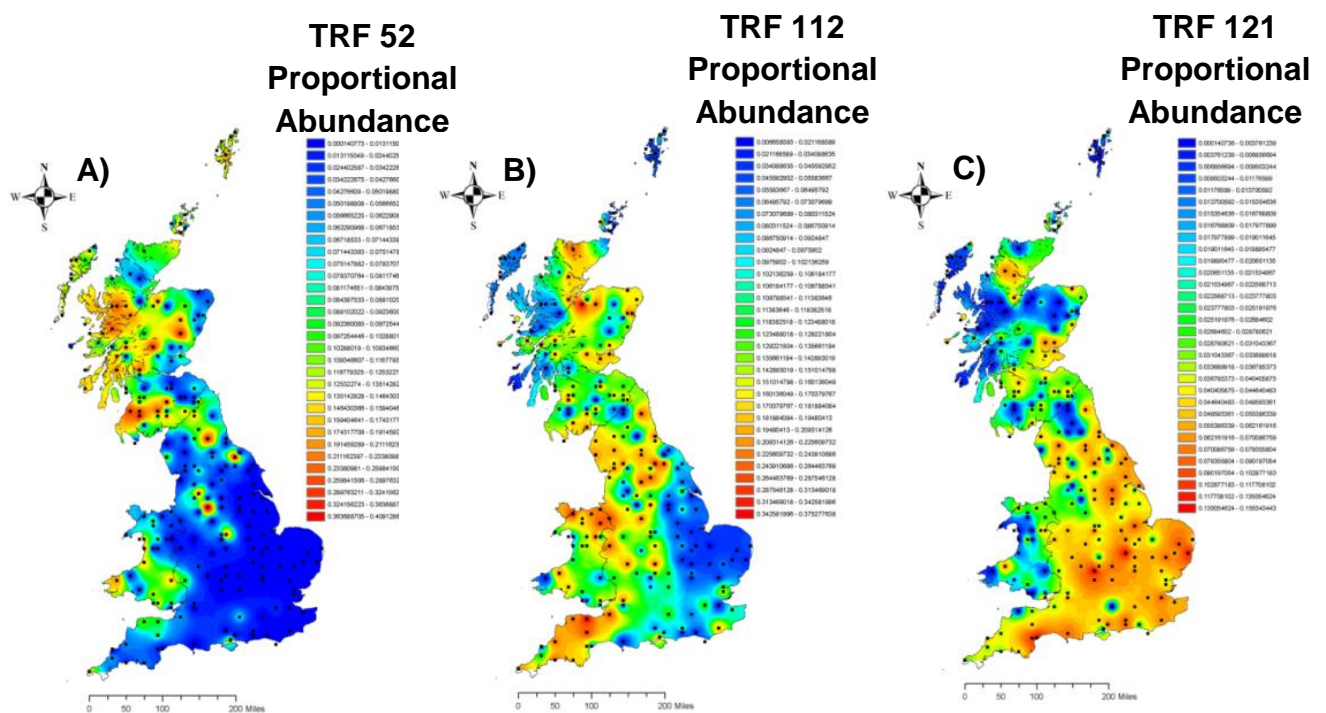


Figure 16: IDW interpolation of the proportional abundance of TRFs A) 52 B) 112 and C) 121 across the UK showing spatially distinct areas of bacterial lineages producing TRFs of these lengths.

3.3.3 Multivariate Recursive Partitioning Analysis of Soil Bacterial Communities.

A multivariate regression tree was created to describe the taxa-environment relationship for soil bacterial communities across the UK (Figure 17). The first split of the data set was based upon the general vegetation classification, with organic soil types supporting heath and bog, moorland grass mosaics and upland wooded communities separating to the right, and typically improved habitats such as crops and weeds, fertile and infertile grasslands, lowland wooded and tall grass and herb communities splitting to the left. Samples originating from areas characteristic of the right hand side of the tree underwent a second split based upon soil pH, with values greater than 5.205 showing a dominance of TRFs 112 (linked earlier to the Alphaproteobacteria) and 52 (Acidobacteria). Highly acidic soils with a pH less than 5.205 were further split based upon C:N ratios with high carbon content soils dominated by TRF 52. Those with higher nitrogen contents had high proportional abundances of TRF 52 and to a lesser degree 112. To the left of the tree, samples originating from typically improved habitats were also split upon the basis of soil pH. More acidic soils were highly dominated by TRF 112 and had low abundances of TRF 52, while higher pH soils exhibited the lowest values associated with TRF 52 but exhibited higher abundances of TRF 112, although these were also in low abundance compared with all other groups. In summary, low pH unimproved soils with high carbon contents were dominated by TRF 52, representing the Acidobacteria, while higher pH habitats primarily contained Alphaproteobacteria. The shifts in the community composition between groups were highly dependent upon the variation in proportional abundance of these two TRFs. TRFs around 121 bp in length (representative of the Actinobacteria) were in comparatively low abundances across all samples, but showed a higher proportional abundance in soils with low C:N ratios. No TRFs dominated the group representing the highest pH ranges (pH > 6.905), and thus a more uniform distribution of bacterial taxa was observed. This suggests that community structure was more even within the neutral and alkaline soils than within groups representing highly acidic soils. This corroborates findings presented earlier showing an increase in alpha diversity and a decrease in beta diversity with increasing soil pH.

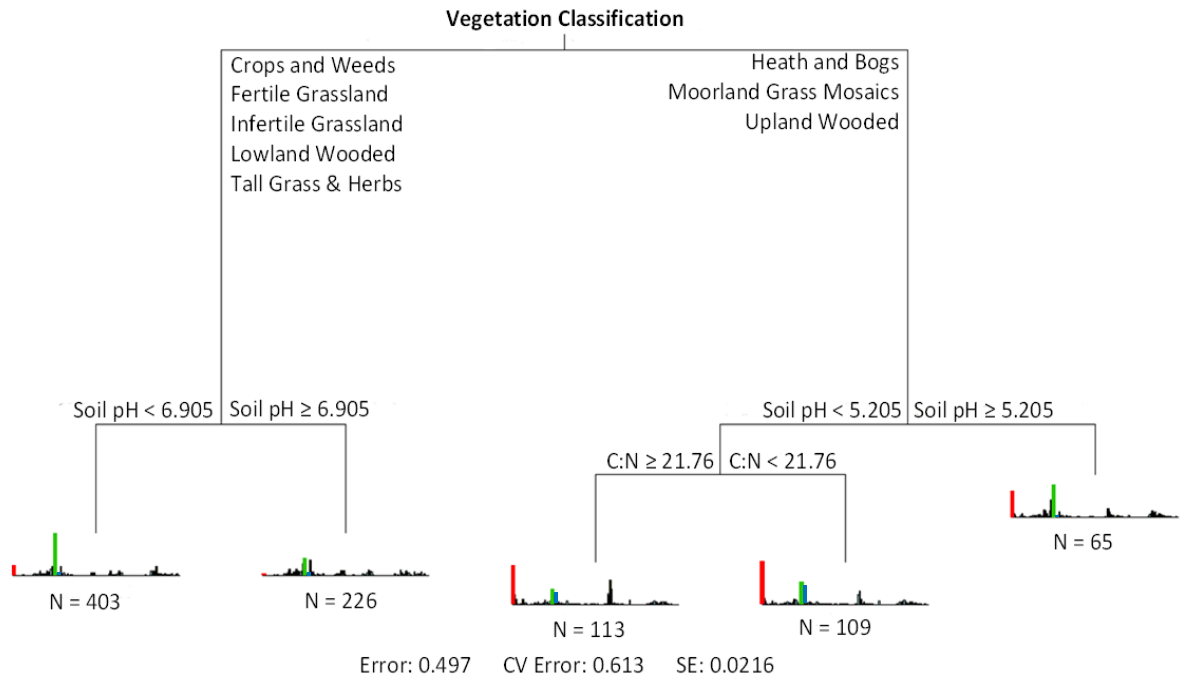


Figure 17: Multivariate regression partitioning tree showing a basic predictive model describing the relationship between environmental conditions and soil bacterial community structure. Successive partitioning of t-RFLP profiles was achieved by minimizing community dissimilarity between t-RFLP profiles within a range of environmental conditions. Each terminal node displays an average t-RFLP profile for samples with specific environmental conditions and the model accounts for approximately 50 % of the variance within the data. Within the terminal bar charts, TRF 52 is highlighted in red, TRF 112 in green, and TRF 121 in blue. Re-calculated from Griffiths *et al.* (2011) with additional environmental variables.

3.4 Discussion

In broad concurrence with several studies examining bacterial community structure in soils (Fierer & Jackson, 2006; Lauber *et al.* 2009; Chu *et al.* 2010), pH was found to be the most important determinant of community composition and diversity over all other environmental variables. In contrast to studies such as that examining soil bacterial diversity and community composition at the landscape scale in America (Fierer & Jackson, 2006), numerous other environmental variables were also identified as important factors. This suggests that soil bacterial communities are structured by similar mechanisms to larger terrestrial organisms, rather than just a single or reduced number of variables. The increase in significance of other environmental variables may have been due to the larger sample size and frequency within this study, allowing a more detailed statistical insight into bacterial community variation over shorter environmental gradients.

3.4.1 Effect of Soil pH on Bacterial Diversity

Alpha diversity showed a typical increase with soil pH. However, in contrast to other large scale surveys of bacterial biogeography (Fierer & Jackson, 2006; Lauber *et al.* 2009) a unimodal distribution was not observed. This, in part, may be because there are few commonly occurring soil types with pH values above 8.5. With the potentially large environmental stresses imposed upon organisms residing within extremely alkaline soils (Sylvia *et al.* 2005), it is possible that alpha diversity would again decrease at higher pH levels to form a unimodal distribution. Soils with extremely high pH are rare globally (IGBP-DIS, 1998), and this is reflected in the sampling design. Within the Americas studies, which observed a decrease in alpha diversity as soils became alkaline across similar pH ranges to those examined here, samples from arid areas in South America may have been responsible for the decline in diversity observed and explain discrepancies between the two studies. Long term arid conditions are not generally present within the UK due to its temperate maritime climate (Peel *et al.* 2007).

Beta diversity is not commonly examined within bacterial community studies despite its importance in ecological systems (Lozupone *et al.* 2007). Due to the higher level of heterogeneity associated with low pH, organic soils, a larger number of potential niches may be available to bacterial taxa when all low pH samples are taken into consideration. Therefore, two low pH samples may each contain relatively few bacterial taxa due to stresses imposed by acidic conditions (and thus produce low alpha diversity measures), but each sample is different in terms of the environmental conditions present, and therefore harbours different communities (thus exhibiting a high beta diversity score). In contrast, higher pH soils impose lower stresses upon the bacterial communities present, and thus alpha diversity is high. However, each high pH soil is similar to every other high pH soil and therefore, at the taxonomic resolution analysed here, each high pH community is similar to every other high pH community, resulting in a low beta diversity measure.

In summation, each individual high pH soil is highly diverse in terms of richness and evenness of bacterial taxa present, but each high pH soil contains a similar bacterial community. Low pH soils individually contain fewer bacterial taxa, but variability in bacterial community structure between low pH samples is greater,

resulting in a larger beta diversity. Meta-communities from low pH environments, representing the combined analysis of communities from multiple low pH soils, were shown to maintain a lower total diversity (gamma) than high pH communities, so that the overall trend of increasing diversity with pH was maintained. Similar trends in soil bacterial biogeography have been observed in terrestrial biomes with distinctly different conditions. A reduction in bacterial alpha diversity with decreasing pH was observed in pyrosequencing assessments of arctic tundra soils, and when these samples were analysed in conjunction with samples obtained from lower latitudes this relationship was maintained (Chu *et al.* 2010). This suggests that the pH effect upon bacterial diversity is common to many different soil ecosystems. However, the discrepancy between studies proposing a unimodal and those proposing a linear distribution of bacterial diversity across soil pH gradients means that there is a need for larger scale, global studies examining bacterial biogeography, or collaborative meta-analysis of landscape scale studies with the same eventual goal in mind.

In several landscape scale studies of soil bacterial biogeography, the emphasis has been mainly upon soil pH. Although the change in bacterial taxa with changing soil pH seemed to primarily define population diversity and community structure within this study, it is proposed that the actual rules governing changes in community composition across ecological gradients are defined by a more complex system of co-correlating environmental variables.

3.4.2 Environmental Influences on Soil Bacterial Community Structure

Although NMDS analysis showed that bacterial community structure had the strongest relationship to soil pH over all other environmental variables, a strong relationship with plant communities was also observed. This shows a definite relationship between above ground and below ground communities, showing that land use (in particular agricultural practices) also plays a part in structuring bacterial communities. It could be argued that this effect is an artefact as a result of pH influencing plant community structure, but inputs into the soil from plants (for example plant root exudate compounds) have been shown to alter local soil environmental conditions, including pH (Grinstead *et al.* 1982;

Dinkelaker *et al.* 2006). It is likely that this local alteration may result in an observable effect upon bacterial community structure at the landscape scale.

Numerous environmental variables co-correlate within the soil habitat, many with soil pH (Rousk *et al.* 2010). For example soil nutrient status is directly linked to soil pH as nitrification rates in low pH soils are reduced due to chloride inhibition (Roseburg *et al.* 1985; Ste-Marie & Pare, 1999). This in turn may be the cause of changes in the community structure of the microbes present within these soil types, even though soil pH may appear to be the direct cause as it is the measured variable. Furthermore, certain soil types become acidic due to leaching of base cations (Rengel, 2003), which is controlled by a number of climatic and edaphic variables such as moisture content, precipitation rates, and temperature (Rengel, 2003). Moisture content is predominantly determined by the rate of soil drainage, which is controlled by soil particle size (Jenny, 1941), while precipitation rates and temperature are a result of latitude and geography. Therefore, although soil pH appears to be the predominant driver of soil bacterial diversity and community structure, many other environmental variables affect, and may also be affected by, changes in soil pH. It is possible that variables which are co-correlated with soil pH, rather than soil pH itself, may be the actual cause of observed differences in bacterial diversity and community structure between different soil samples. However, partitioning the variance in bacterial community structure attributed to numerous co-correlating environmental and edaphic variables, both measured and unmeasured, is fundamentally difficult. This is not necessarily a problem as it reflects the complex conditions found within actual soil ecosystems rather than experimentally manipulated soils. Measuring large numbers of environmental and edaphic properties, as was performed here, allows a simplified but holistic description of the most important variables determining bacterial community structure and diversity to be produced.

T-RFLP profiles and multivariate analyses suggest that a few dominant taxa were responsible for the large biogeographic effects observed in soil bacterial communities. *In-silico* digestion of a global 16S rRNA gene database showed that restriction fragments identified as important indicator OTUs between different soils represent common soil bacteria. Alphaproteobacteria and Actinobacteria have been well documented as key groups of organisms within

soil ecosystems (Lauber *et al.* 2009; Janssen, 2006), known to contain particular groups with a degree of plant specificity (Costa *et al.* 2006). TRFs representing both Alphaproteobacteria and Actinobacteria had, in general, low abundances in acidic soils, and higher abundances in neutral and alkaline soils. Acidobacterial taxa showed a sharp decline with increasing pH, and were deemed characteristic of low pH environments. Despite the overall decline in Acidobacteria in soils of higher pH, particular subgroups of the Acidobacteria have been shown to have contrasting relationships with soil pH. Groups 1 and 2 have been shown to be negatively correlated with soil pH, whilst group 6 Acidobacteria were found to increase in abundance with increasing soil pH (Sait *et al.* 2006; Jones *et al.* 2009). This implies that pH also has an effect upon bacterial community structure at taxonomic resolutions greater than t-RFLP can provide. The use of t-RFLP for community profiling of soil bacteria at the landscape scale only shows broad scale changes in bacterial groups over environmental gradients, due to the low taxonomic resolution inherent to the method (Fierer, 2007). The identification of these broad scale patterns allows generalised predictions to be made of abundances of broad taxonomic groups in a given set of environmental conditions. However, it is unclear whether these patterns hold true at taxonomic resolutions closer to the species level, or for the rarer members of a community.

3.4.3 Mapping Bacterial Communities

In an effort to simplify the effects of co-correlating environmental variables upon soil bacterial community structure, interpolated mapping of ordination scores allows observations to be made concerning geographical differences in bacterial populations. Here a generalised “north south divide” in bacterial community structure is seen, with communities located towards the left of the NMDS ordination predominantly found in the low pH organic areas of west Scotland, and those to the right of the ordination in higher pH soils i.e. those used for agriculture in the south east England. *A priori* knowledge of typical climatic conditions allows for geographical links to be made between areas of similar bacterial community structure. For example, acidic upland soil types from west Scotland and parts of south west Wales support similar plant species and exhibit a similar bacterial community structure. Overarching geological and environmental conditions that cause these low pH organic soils to form are also

similar in nature i.e. high rainfall, and colder, typically harsh climates. Conversely, soils in milder climates found towards the agriculturally dominated eastern side of Scotland are more similar in bacterial community structure to those of the arable south of the UK.

The interpolation of ordinations describing above ground plant community variability also provides a visual indication of the link between above ground and below ground biomes. Areas with distinct floral communities generally exhibited distinct bacterial communities; however, these floristic communities may be largely the result of co-correlating environmental variables acting upon plant community structure. It is likely that the links between above ground and below ground communities shown here are important to a degree in reciprocal structuring of both above ground and below ground ecological systems.

Mapping the proportional abundance of individual TRFs across the UK shows how particular groups of bacterial taxa respond to environmental variability. TRF 52, identified as predominantly acidobacterial taxa, are in high abundance within areas highlighted as containing a typical acidic bacterial community and were thus identified as indicative of soil types with properties outlined above. Alphaproteobacteria and Actinobacteria were generally in low abundances where Acidobacteria dominated, and were generally more common in areas with bacterial communities associated with the right of the NMDS ordination. Actinobacteria are known to be indicative of good soil health (van Dijk, 2008) and were found predominantly in areas known for their fertility and use in agricultural crop production.

3.4.4 Conclusions

In conclusion, broad scale differences in soil bacterial groups are predominantly controlled by environmental differences, particularly soil pH. Dominant plant species and other co-correlating environmental variables are important but to a lesser degree. There was no evidence to suggest a unimodal distribution of soil bacterial diversity with pH; instead, a linear increase was observed. Beta diversity decreased with increasing pH exemplifying greater between community variation in soils of low pH. Mapping of bacterial community structure and the abundances of particular bacterial groups provided a visual

representation of the relationship between soil bacterial communities and climatic gradients.

Here general patterns in soil bacterial biogeography have been elucidated across the UK, representing patterns at the landscape scale. However, the use of t-RFLP did not allow the comparison of bacterial community composition through direct taxonomy assignments between soils of different edaphic and environmental conditions. The application of next generation sequencing strategies allows a more detailed interrogation of UK soil bacterial communities.

3.4.5 Future Work

The biogeography of soil bacterial communities is likely to be somewhat different at finer spatial resolutions. At the microbe scale, the physical separation of soil bacteria within surface soil microhabitats has given rise to the theory that competition between soil bacterial communities is reduced, or acts in a different way from that of plants and animals, due to the infrequent direct interaction between soil bacterial organisms (Zhou *et al.* 1997; Tiedje *et al.* 2001). Interaction between soil bacterial organisms is thought to be relatively infrequent, and limited to short periods of time after soil saturation (Young & Ritz, 2000). After soil saturation, the proliferation and dominance of specific microbial lineages has been noted, suggesting competition rates are regulated by habitat isolation (Tiedje *et al.* 2001). Therefore the biogeography of soil bacterial communities may be governed by entirely different processes at the fine spatial scale compared with the landscape scale. Localised high density soil sampling regimes may provide further insight into the factors affecting microbial biogeography at the fine scale, and help construct an ecological framework defining the environment-taxa relationship across multiple spatial scales. Furthermore this type of study would also provide a greater insight into more complex processes shaping bacterial community structure, for example within and between kingdom competition and predation.

The use of next generation sequencing methods allows the interrogation of soil bacterial communities at levels approaching that of the species (Roesch *et al.* 2007). Coupled with the ability to capture a large proportion of the diversity present within a sample and assign taxonomic classifications to defined operational taxonomic units (OTUs), the use of these sequencing methods will

allow a more detailed insight into the factors affecting soil bacterial biogeography. T-RFLP analysis describes ecological patterns at comparatively low taxonomic resolutions; it is unclear whether these patterns hold true at resolutions closer to that of the species. Further investigation using next generation sequencing methods will highlight the factors affecting soil bacterial diversity and community structure at varying levels of taxonomic resolution, and help define the taxa-environment relationship at multiple taxonomic scales.

Chapter 4: Pyrosequencing Analysis of Soil Bacterial Communities over a Natural pH Gradient: A Comparison of Taxonomic Marker Genes

4.1. Introduction

4.1.1 Background

Defining the ecological characteristics of microbial populations from the environment is a challenging task fraught with potential biases and technical limitations. Soil bacterial populations are commonly regarded as some of the most diverse on the planet (Fierer & Jackson, 2006; Torsvik *et al.* 1990). In order to accurately examine diversity and changes in community composition within these complex systems, a succession of methodologies has arisen over the last decade or so. Culture based methods can be used to identify bacterial species present in a sample based on their phenology, but a high proportion of bacterial species present in a soil sample are non-culturable under standard laboratory conditions (Ferrari *et al.* 2005; Janssen *et al.* 2002). Molecular based methods exploiting the variability in highly conserved genes capture a larger proportion of the bacterial diversity present (Dunbar *et al.* 1999; Sait *et al.* 2002), and thus their use has become a standard approach in microbial studies. Numerous studies have utilised different approaches in an attempt to examine soil bacterial communities from an ecological standpoint, ranging from rapid “fingerprinting” methods such as t-RFLP (Fierer & Jackson, 2006) and DGGE (Nakatsu *et al.* 2000; Muyzer & Smalla, 1998), through to sequencing methods such as Sanger sequencing (Thomson *et al.* 2010) or contemporary pyrosequencing approaches (Lauber *et al.* 2009; Metzker, 2010).

Pyrosequencing relies upon analysis of DNA sequences, which can then be identified to particular taxa, and binned into representative operational taxonomic units (OTUs) to generate an assessment of community composition and diversity. In order to assess the diversity of complex microbial populations using this method, highly conserved genes must be sequenced. The 16S rRNA gene has become the taxonomic marker gene of choice due to the extensive research carried out upon its structure (Gutell, 1993; Neefs *et al.* 1990; Noller & Woese, 1981), and the availability of large open access sequence repositories (DeSantis *et al.* 2006; Altschul *et al.* 1990; Pruesse *et al.* 2007). The 16S rRNA

gene is highly conserved across all bacterial taxa, but contains 9 regions of hypervariability, known as V regions (Neefs *et al.* 1990). This inherent variability makes it possible to distinguish between different bacterial taxa. Limits in the sequence length achievable through pyrosequencing mean that bacterial community profiling studies utilising this technology target short lengths of the 16S rRNA gene (Kim *et al.* 2011). Until recently, most studies have examined single hypervariable regions (Roesch *et al.* 2007; Huse *et al.* 2008; Liu *et al.* 2007). However, the increase in sequence length possible with new pyrosequencing methods allows analysis of several contiguous hypervariable regions, for example the V1 – V3 regions (ca. 470 bp *E.coli* numbering 28 to 519) or the V6 – V9 regions (ca. 600 bp *E.coli* numbering 939 to 1492).

Soil pH is now known to be one of the key drivers of soil bacterial diversity and community structure. Landscape scale surveys have been carried out using fingerprinting methods (Chapter 3), small scale clone library analysis (Griffiths *et al.* 2011), and pyrosequencing analysis (Fierer *et al.* 2007). One of the consistent findings of all these studies is that bacterial diversity within low pH environments is significantly less than that of higher pH soils. However particular studies have observed different diversity relationships in high pH soils. Several studies found a linear increase in soil bacterial diversity with pH (Rousk *et al.* 2009; Hartman *et al.* 2008; Griffiths *et al.* 2011), whilst others identified a unimodal relationship, with a reduced diversity seen at high pH extremes (Fierer & Jackson, 2006; Chu *et al.* 2010). The hypervariable regions analysed in these studies were not consistent, and thus a direct comparison of findings cannot be made. Furthermore it is unclear whether the biogeographic patterns highlighted in Chapter 3 are consistent across multiple taxonomic levels.

Using DGGE and *in-silico* analysis of Ribosomal Database Project (RDP) sequences, Yu & Morrison (2004) showed that there was a distinct difference in community profiles when different hypervariable regions were targeted (Yu & Morrison, 2004). Furthermore, the sequence variability between hypervariable regions across 218 aligned reads was significantly different. V1 – V3 regions were shown to be the most divergent, while V6 – V9 were less divergent but still maintained a higher divergence than all other V regions. It is therefore likely that ecological community descriptors calculated for complex communities, such as

those of soil bacteria, may vary with the choice of amplified 16S rRNA region. Therefore further work is required to compare and contrast the effect of variable region choice when complex bacterial communities across environmental gradients are examined by pyrosequencing.

4.1.2 Aims

As the previous chapter has identified patterns of bacterial biodiversity in relation to variation in environmental parameters, principally soil pH, using a method with low taxonomic resolution, the primary aim of this chapter is to investigate the responses of bacterial populations to soil pH at a high taxonomic resolution using a taxonomically informative deep sequencing method to determine if soils of similar physicochemical properties are truly dominated by the same taxa, or if this finding is related to the taxonomic resolution of the analytical method used. Because of the known effects of primer choice upon examinations of soil bacterial communities, the secondary objective of this study is to assess soil bacterial diversity and community composition using pyrosequencing assays targeting the V1 - V3 and V6 – V9 regions of the 16S rRNA encoding gene.

4.2. Methods

4.2.1. Sample Selection

From the Countryside Survey samples described in section 2.2.1 a subset of 15 soil cores was selected from low pH (pH 4.3 +/- 0.23; n = 5), medium pH (pH 6.15 +/- 0.08; n = 5) and high pH (pH 8.28 +/- 0.16; n = 5) environments. Cores were taken from geographically independent areas (Appendix B), representing low, medium and high pH soil types across the UK.

4.2.2. Nucleic Acids Extraction Procedure

For each soil sample, three repeated nucleic acid extractions were performed on three replicates. DNA was extracted as described in section 2.2.2, and then pooled in equal concentrations.

4.2.3 Massively Parallel Bacterial Tag Encoded FLX Amplicon Pyrosequencing (BTEFAP)

The bTEFAP sequencing method was performed by Research and Testing Laboratory (Lubbock, TX, USA). Nucleic acid concentration was adjusted to 100 ng μl^{-1} and 1 μl was used as a template for a single stage PCR amplification of 30 cycles utilizing HotStar Taq Plus Master Mix Kit (Qiagen, Valencia, CA, USA). Two fusion primer sets were chosen to amplify two distinct regions of the 16S rRNA gene commonly used for bacterial community profiling studies. The V1 – V3 region (ca. 500 bp) was amplified using primers 28F (5'- GAG TTT GAT YMT GGC TC -3'), and 519R (5'-GYT ACC TTG TTA CGA CTT -3')(Lane *et al.* 1991), while the V6 – V9 region (ca. 600 bp) was amplified using primers 939F (5'- TGA AAG AAG AGG GTC GGC AGA TAA -3') and 1492R (5'- GYT ACC TTG TTA CGA CTT – 3') (Zhao *et al.* 2011). Primer pairs contained a specific Linker A Tag on the forward primer and a Linker B Tag on the reverse primer. All samples were identified after sequencing via unique DNA sequence barcodes. PCR reactions were carried out under conditions optimised by the Lubbock Research and Testing Laboratory. After amplification, all samples were pooled in equal concentrations and purified using Agencourt Ampure Beads (Agencourt Bioscience Corporation, MA, USA).

Prior to sequencing, DNA quality control of amplicons was checked using Bio-Rad Experion Automated Electrophoresis Station (Bio-Rad Laboratories, CA, USA) and a TBS-380 Fluorometer (Turner Biosystems, CA, USA). Quality checked amplicons were combined with DNA capture beads in an approximate ratio of 1 amplicon to 1 bead, to ensure each bead presented a single amplicon. Beads underwent oil immersion PCR to saturate each bead with amplicon copies. Post amplification, beads underwent purification and recovery steps as outlined in the manufacturer's instructions, then amplicons were denatured with NaOH and sequencing primers were annealed. Sequencing reactions were performed using the Roche 454 FLX instrument with Titanium reagents (Roche, CT, USA) according to optimised procedures used by the Research and Testing Laboratory. A single half of the plate was used in conjunction with 69 other unrelated samples identified and excluded from subsequent analysis by the use of unique DNA barcode sequences.

4.2.4. Sequence Processing and Data Analysis

Qiime v1.3.0 (Caporaso *et al.* 2010) was used in a virtual box format to perform all bioinformatics steps. Prior to sequence analysis, libraries were quality checked and trimmed accordingly. Concatenated FASTA files and the corresponding quality file were used to screen libraries and select for sequences with a minimum length of 200 bp, and a minimum mean quality score of 30. Additionally, sequences with any ambiguous bases were removed, along with those with a maximum homopolymer length longer than 6 bp. The primer mismatch allowance was set at 2, and a quality window of 50 bp was used to trim bad quality regions at the first base of the window where the mean quality score of the window fell below 30. Barcodes and primers were also trimmed from all sequences.

To assess the effects of environmental variables upon bacterial community dissimilarity using either set of variable regions, site by OTU matrices generated in Qiime at 97 % OTU sequence identity were analysed by NMDS using the metaMDS function in R. Environmental variables were fitted to the ordination using the envfit function in vegan.

To compare sequence dissimilarity between the V1 - V3 and V6 - V9 regions, two high quality alignments were created from a subset of quality checked sequences with a minimum length of 400 bp ($n = 1000$ per subset) using MUSCLE (Edgar, 2004). Iterations of the alignment ran until convergence was achieved. A pairwise Euclidean distance matrix was generated for each alignment and total sequence variability was compared between the two regions using boxplots and Welsch's two sample T-test.

Uclust (Edgar, 2010) was used to cluster all sequences into operational taxonomic units (OTUs) under default settings. Seed sequences were generated and formed clusters based upon 5 levels of sequence similarity: 97 %, 94 %, 90 %, 85 % and 80 %. Sequences that did not match a seed at the defined level of sequence similarity became new seeds for further OTU clusters. The seed sequence within an OTU cluster was used as the representative sequence for each OTU, and taxa assignments were given to each sequence using the RDP naïve Bayesian classifier (Wang *et al.* 2007) with a confidence value of 0.8. Representative sequences were aligned using PyNast (Caporaso

et al. 2010) against the aligned version of the bacterial 16S rRNA core Greengenes Database (n = 10,000) (DeSantis *et al.* 2006) with a minimum percentage sequence identity to the database of 0.75. Chimeric sequences were detected using ChimeraSlayer (Haas *et al.* 2011) and removed from subsequent analysis steps. Alignments were filtered for vertical gaps and chimeric sequences, and a lane mask filter was applied prior to constructing phylogenetic trees with FastTree (Price *et al.* 2010). Rarefied alpha diversity metrics were also computed using the Qiime pipeline, and relative abundances were calculated at various levels of taxonomic classifications (phylum – genus).

As pyrosequencing results in an uneven distribution of sequences returned per sample, the sequencing effort for each sample may directly influence calculated diversity indices. To account for the differences in sequencing effort per sample, rarefied Chao1 scores were calculated for each sample to give an indication of relative diversity (Chao, 1984). Chao1 scores were calculated using the equation:-

$$S_1 = S_{obs} + \frac{F_1^2}{2 F_2}$$

Where S_{obs} is the number of OTUs in a subsample, F_1 is the number of singletons (i.e. OTUs that only occur once in the rarefied subsample), and F_2 is the number of OTUs with exactly 2 occurrences in the rarefied subsample.

Phylogenetic trees were visualised using ITOL (Letunic & Bork, 2011) and the mean proportional abundance of each OTU within a pH class was plotted in the form of a bar graph at the terminal node.

Principal coordinate (PCo) analysis was used to visualise jackknifed (n = 10) unweighted Unifrac (Lozupone & Knight, 2005) scores for each sample as a phylogenetically based assessment of community dissimilarity. Hierarchical clustering using an unweighted pair group method with arithmetic mean (UPGMA) was also implemented to visualise the clustering of samples within different pH classes. Jackknife support was utilised by multiple comparisons (n = 10) of UPGMA trees constructed from a subset of sequences (75 % of the smallest sample) with the tree created from all sequences. Nodes of the

resultant tree were then coloured to show Jackknife support (red = 75 – 100 %, yellow = 50 – 75 % and green 25 – 50 %).

The software package R (Ihaka & Gentleman, 1996) was used to visualise data generated in Qiime in the form of line or bar graphs using the packages Vegan (Oksanen *et al.* 2011); BioDiversityR (Kindt & Coe, 2005); and RcolorBrewer (Neuwirth, 2011). Phylogenetic trees were visualised using ITOL (Letunic & Bork, 2011) and the mean proportional abundance of each OTU within a pH class was plotted in the form of a bar graph at the terminal node.

To calculate the taxa most responsible for the community dissimilarity observed between the different pH groups, indicator species analysis was performed using the *indval* function in the *labDSV* package in R (Roberts, 2010), according to the principles outlined in Dufrene & Legendre (1996).

4.3 Results

4.3.1 Sequence Processing

Tables 4 A) and B) show the number of sequences removed at each stage of filtering for the V1 – V3 region and V6 – V9 region analysis per sample, and tables 4 C) and D) show the numbers of OTUs returned in total, and per sample, at differing levels of OTU clustering for the V1 – V3 and V6 – V9 analysis. All data presented in the main text is at 97 % sequence similarity unless otherwise stated in the text.

A)	Sequences post filtering stage	Low pH samples 1 - 5				
	Pre filtering	19545	22874	26001	13188	19745
	Quality	14509	19378	9658	14814	17238
	Length	12793	17166	8425	13090	15275
	Homopolymer	12721	17160	8375	13026	15208
	Primer Mismatch	5211	13791	3896	7648	9382
	Chimeras	5046	13126	3536	7118	8037
		Medium pH Samples 1 - 5				
	Pre filtering	16709	17721	13192	20619	12624
	Quality	12328	13449	9496	15832	9447
	Length	10865	11155	7955	14140	7137
	Homopolymer	10806	11064	7884	14089	7082
	Primer Mismatch	4262	5858	6062	11657	7082
	Chimeras	3479	2591	4453	9945	3906
		High pH Samples 1 – 5				
	Pre filtering	17783	14699	21303	16481	7650
	Quality	13044	10716	15039	11761	5544
	Length	11593	9375	12625	10237	4700
	Homopolymer	11513	9315	12559	10184	4669
	Primer Mismatch	4277	4848	7412	5182	4669
	Chimeras	3970	4486	6880	4750	3858
B)	Sequences post filtering stage	Low pH samples 1 - 4				
	Pre filtering	8794	13192	26001	20619	--
	Quality	6325	9469	19378	15832	--
	Length	6301	7955	17166	14140	--
	Homopolymer	6299	7884	17106	14089	--
	Primer Mismatch	6299	2822	3987	2432	--
	Chimeras	6293	2211	3888	2211	--
		Medium pH Samples 1 - 5				
	Pre filtering	19545	14699	21303	16481	13188
	Quality	14509	10716	15039	11761	9658
	Length	12793	9375	12625	10237	8425
	Homopolymer	12721	9315	12559	10184	8375
	Primer Mismatch	8510	5467	6147	6002	5479
	Chimeras	8233	4812	6018	5773	5110
		High pH Samples 1 – 5				
	Pre filtering	16709	17783	22874	17721	19745
	Quality	12328	13044	17238	13449	14814
	Length	10865	11593	15275	11155	13090
	Homopolymer	10806	11513	15208	11064	13026
	Primer Mismatch	7544	8236	6826	6206	6378
	Chimeras	6978	7920	6242	5614	5909

C)	OTU Clustering Similarity (%)	Low pH samples 1 – 5				
	97	692	852	854	513	836
	94	464	540	566	346	558
	90	308	361	393	245	373
	85	212	238	261	164	253
	80	137	164	177	127	179
		Medium pH Samples 1 – 5				
	97	817	576	863	702	746
	94	579	423	623	509	535
	90	403	299	415	352	380
	85	266	204	266	239	249
	80	170	133	166	154	158
		High pH Samples 1 – 5				
	97	787	944	1305	931	794
	94	588	673	893	658	599
	90	402	469	549	432	405
	85	275	317	362	271	267
	80	181	198	209	166	173
D)	OTU Clustering Similarity (%)	Low pH samples 1 - 5				
	97	642	315	466	423	--
	94	401	209	295	275	--
	90	221	133	156	177	--
	85	123	88	96	100	--
	80	54	42	44	49	--
		Medium pH Samples 1 – 5				
	97	973	749	878	689	780
	94	551	428	523	444	459
	90	294	228	274	239	269
	85	134	124	149	129	141
	80	53	56	56	52	55
		High pH Samples 1 – 5				
	97	844	690	825	657	861
	94	434	365	416	344	498
	90	243	198	221	183	277
	85	110	104	110	105	145
	80	49	47	48	47	55

Table 4 A) Shows the number of sequences returned post filtering for the V1 – V3 region analysis while Table 4 B) shows the number of sequences returned post filtering for the v6 – v9 region analysis. Table 4 C) shows the number of OTUs generated at multiple levels of OTU clustering for the V1 – V3 region analysis, while table 4 D) shows this information for the V6 – V9 region analysis. Total number of OTUs generated for the V1 – V3 region analysis at differing levels of OTU clustering was: 97 % = 7797, 94 % = 4274, 90 % = 2195, 85 % = 1090, 80 % = 534. Total number of OTUs generated for the V6 – V9 region analysis at differing levels of OTU clustering was: 97 % = 5089, 94 % = 2291, 90 % = 877, 85 % = 288, 80 % = 85.

4.3.2 Effects of Environmental Variables upon Bacterial Community Dissimilarity

NMDS ordinations of the site by OTU matrices generated in Qiime were constructed, and environmental variables were fitted to the plot in the form of a least squares regression for both sets of variable regions analysed. Significant environmental factors were added to the plot in the form of vectors of which the length and orientation indicate the magnitude and direction of the effect (Figure 18-19).

Using either variable region to analyse the effects of environmental variables upon bacterial community dissimilarity, soil pH, plant DCA axis 1 scores, soil moisture, and loss on ignition were all significant factors. Soil pH had the strongest effect upon the separation of the data points in both instances. Sequences spanning the V1 – V3 region showed a more distinct pH effect with low, medium, and high pH samples exhibiting tight within group clustering. Within the analysis of sequences spanning the V6 – V9 region, low pH samples separated from high and medium pH samples, but a difference between the community structure of medium and high pH samples was not apparent.

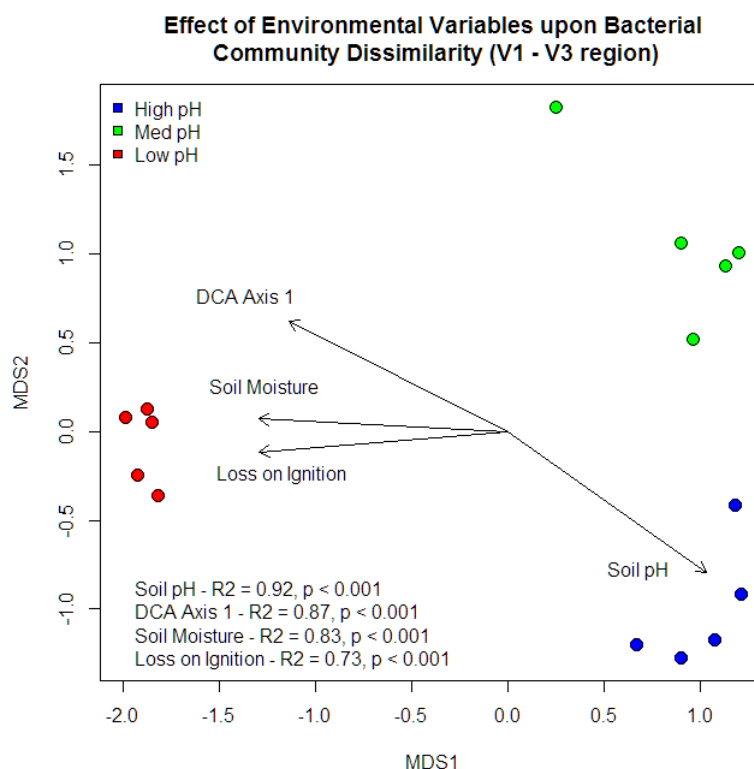


Figure 18: NMDS ordination showing that the community dissimilarity of soil bacterial populations is largely governed by soil pH, plant cover (DCA Axis 1), soil moisture, and loss on ignition when OTUs at 97 % sequence dissimilarity are examined using V1 – V3 region pyrosequencing analysis. Stress value = 0.11.

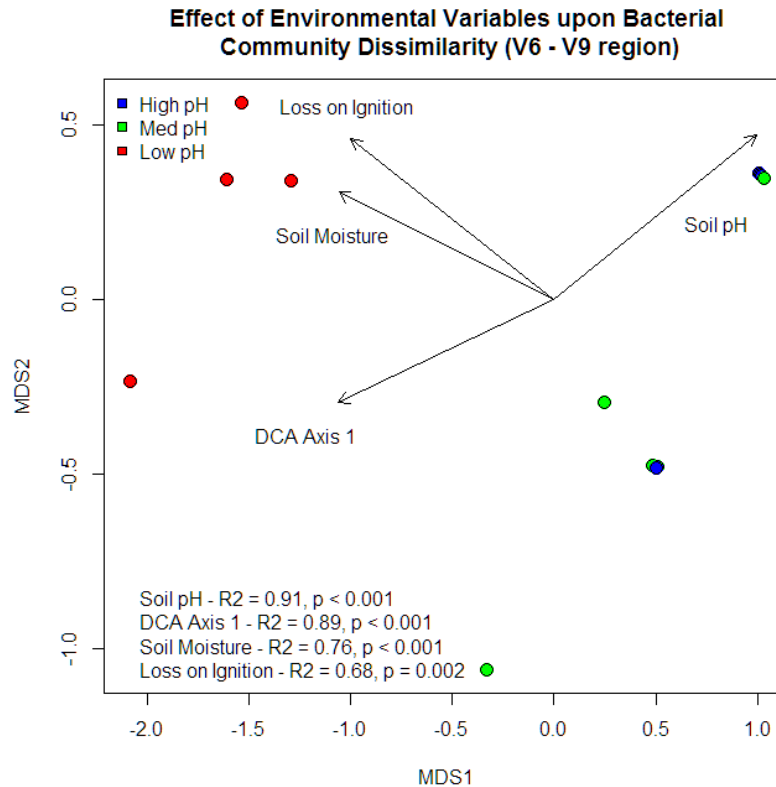


Figure 19: NMDS ordination showing that the community dissimilarity of soil bacterial populations is largely governed by soil pH, plant cover (DCA Axis 1), soil moisture, and loss on ignition when OTUs at 97 % sequence dissimilarity are examined using V6 – V9 region pyrosequencing analysis. Stress value = 0.17.

4.3.3 Dissimilarity of the V1 – V3 and V6 – V9 Regions of the 16S rRNA Gene

As a preliminary comparison of the two contiguous regions of hyper variability, 1000 high quality sequences were picked at random from each set of sequences (V1 – V3 and V6 – V9 regions) equally distributed across all samples. Pairwise comparisons of sequences were performed and the mean sequence dissimilarity was found to be significantly higher for V1 - V3 sequences than V6 - V9 sequences (Welch's T-test; $p < 0.001$) (Figure 20). This suggests that there is a higher level of sequence variability inherent to the V1 – V3 region, potentially allowing a greater level of discrimination between closely related bacterial taxa. To assess the use of these regions of the 16S rRNA gene for next generation sequencing based community profiling methods, further analysis was undertaken to examine the observed effects of pH upon community dissimilarity, diversity, and structure.

Pairwise DNA Sequence Dissimilarity of the V1 – V3 region and V6 – V9 region of the 16S rRNA Gene (subset n = 1000)

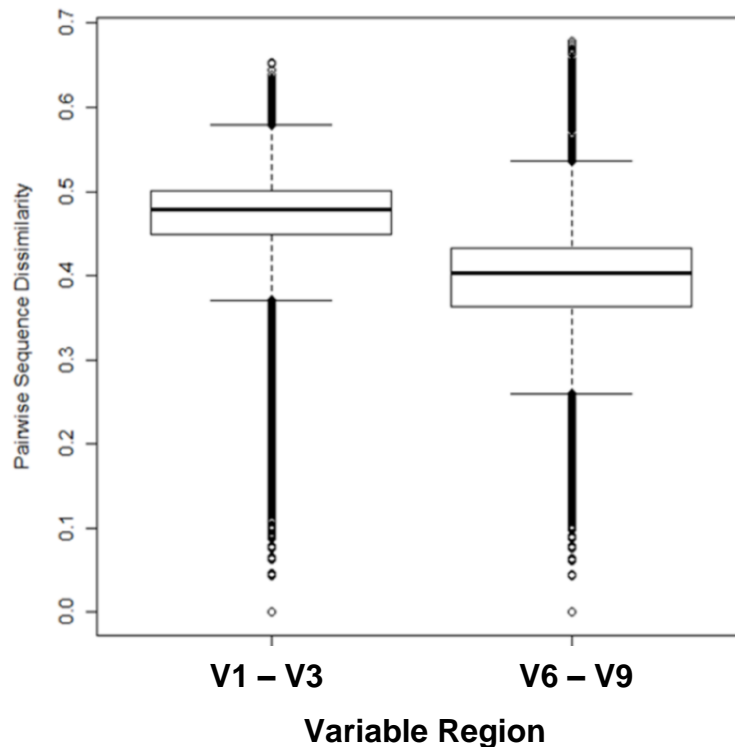


Figure 20: Boxplots showing the mean pairwise dissimilarity for a subset of high quality aligned sequences from the V1 – V3 and V6 – V9 regions of the 16S rRNA gene (n = 1000 per region). The mean pairwise dissimilarity was significantly lower in the V6 – V9 region sequences (Welch's two sample T-test $p < 0.001$) signifying a lower "between taxa" variability.

4.3.4 Phylogenetic Analyses of Community Dissimilarity

Unweighted Unifrac based distance matrices were calculated for both sets of contiguous hypervariable regions. Multiple subsets of the sequences were used for Jackknife support (n = 10). PCoA was implemented to visualise the separation of the samples in a reduced number of dimensions.

Jackknifed Unweighted Unifrac PCO Ordination of Bacterial Communities from Low, Medium, and High pH Soils

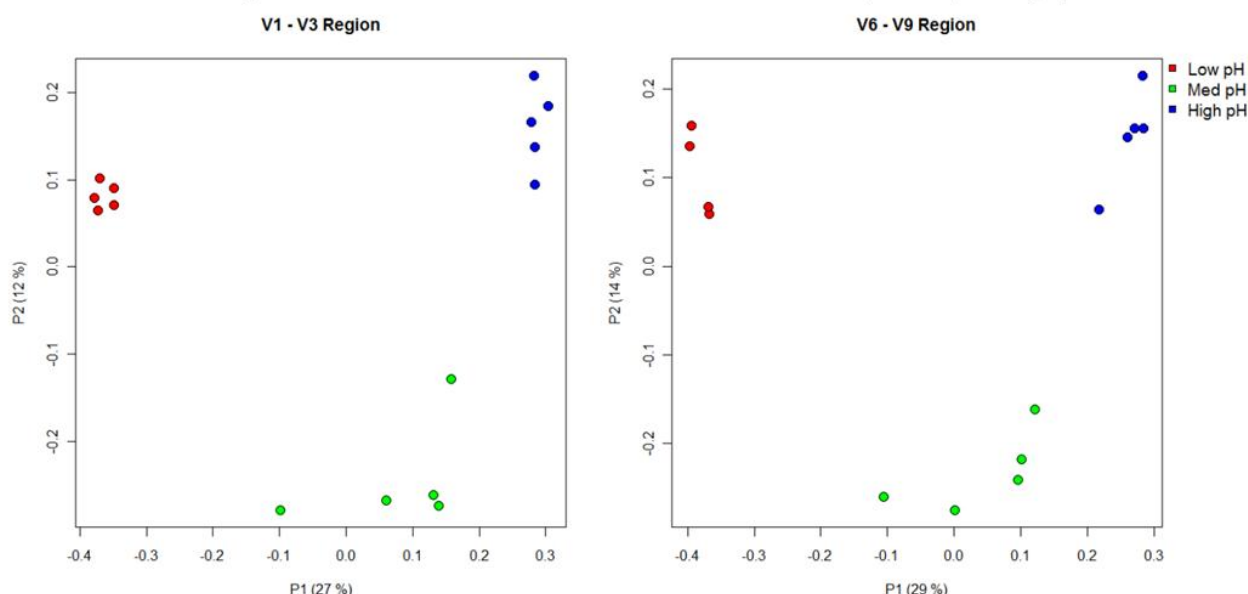


Figure 21: Jackknifed (n = 10) principal co-ordinate (PCo) analysis of un-weighted Unifrac distance matrices indicating an observable effect of pH upon bacterial community structure irrespective of 16S rRNA region targeted. Confidence ellipsoids around each point (defined by Jackknife support) suggest a high degree of confidence in each plotted position as ellipsoid area is non-visible.

Jackknifed UPGMA Trees Showing Community Dissimilarity with Respect to Soil pH.

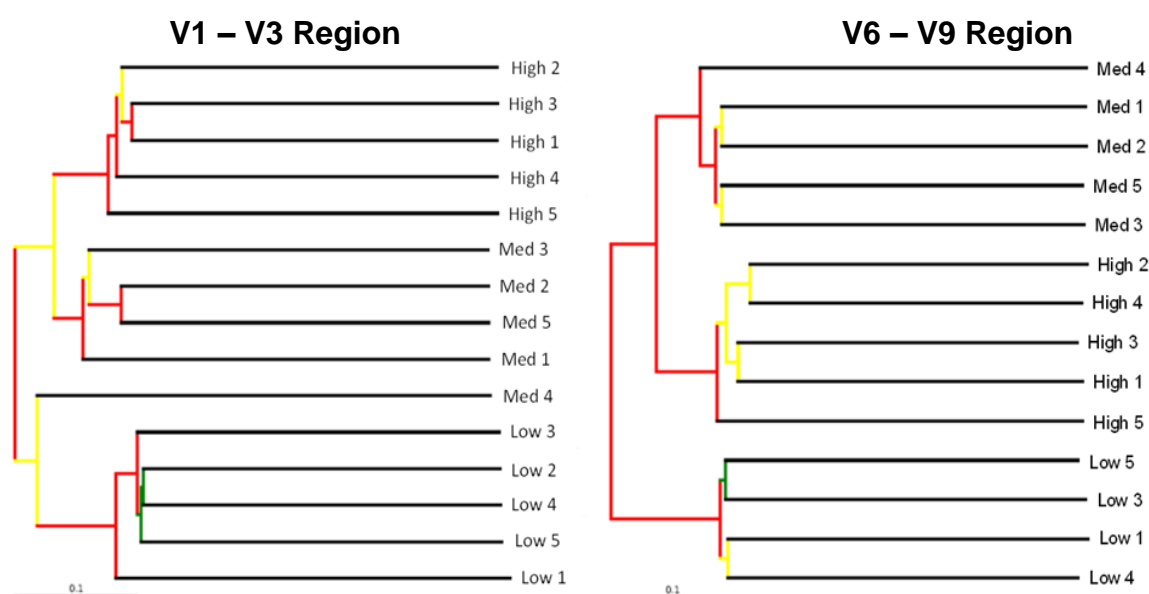


Figure 22: UPGMA hierarchical clustering used to generate Unifrac scores is displayed in the form of trees. Here, not only does within pH group variability remain lower than between pH group variability but medium and high pH samples are shown to be more similar to each other when compared with the low pH samples using either set of hypervariable regions. This is indicated by the first split of the tree on the left hand side with very high Jackknife support. Red = 75 100 %, yellow = 50 – 75% and green = 25 – 50 %.

Jackknife support suggests a high degree of confidence for each plotted point (Figure 21). A distinct separation of the samples based upon pH was observed irrespective of the 16S rRNA gene region targeted. To analyse this effect in further detail the hierarchical clustering used to generate the Unifrac scores was visualised in tree form (Figure 22). One low pH sample did not undergo

successful pyrosequencing and low quality scores dictated that it should be abandoned from subsequent analysis. No attempt at re-sequencing was made due to cost constraints.

Consistent with the jackknifed Unifrac calculations, the within pH class community variability was lower than the between pH class community variability. Here pH had a distinct observable effect upon bacterial community structure, and similar overarching patterns are observed irrespective of the 16S rRNA region targeted for community profiling.

4.3.5 Alpha Diversity

To examine the effect of soil pH upon bacterial diversity at various levels of taxonomic resolution, and to contrast the diversity results generated from V1 – V3 and V6 – V9 region sequences, OTUs were generated at multiple levels of sequence similarity as recommended by Roesch *et al.* (2010). Chao1, a non-parametric diversity estimate (richness) was used to compare alpha diversity between pH groups in a rarefied fashion to avoid discrepancies in sequencing intensity between samples. Lower diversity (richness) scores were seen when low pH soils were compared with high and medium pH soils independent of region targeted at 97 % sequence similarity (Figure 23). Diversity (Chao1 richness) was greater in the V1 – V3 region analysis than the V6 – V9 region analysis for all pH groups. Different patterns in diversity (Chao1 richness) were observed between pH groups depending upon the variable regions used for analysis. V1 – V3 region sequences showed an increase in Chao1 scores with increasing pH at every subset number of sequences analysed, whereas the distinction between medium and high pH samples, in terms of diversity (Chao1 richness), was not as defined.

When Chao1 scores were calculated for OTUs clustered at lower levels of sequence similarity, overall Chao1 diversity was considerably lower in samples where the V6 – V9 regions were targeted compared with Chao1 scores generated from V1 – V3 region sequences (Appendix H i - iv).

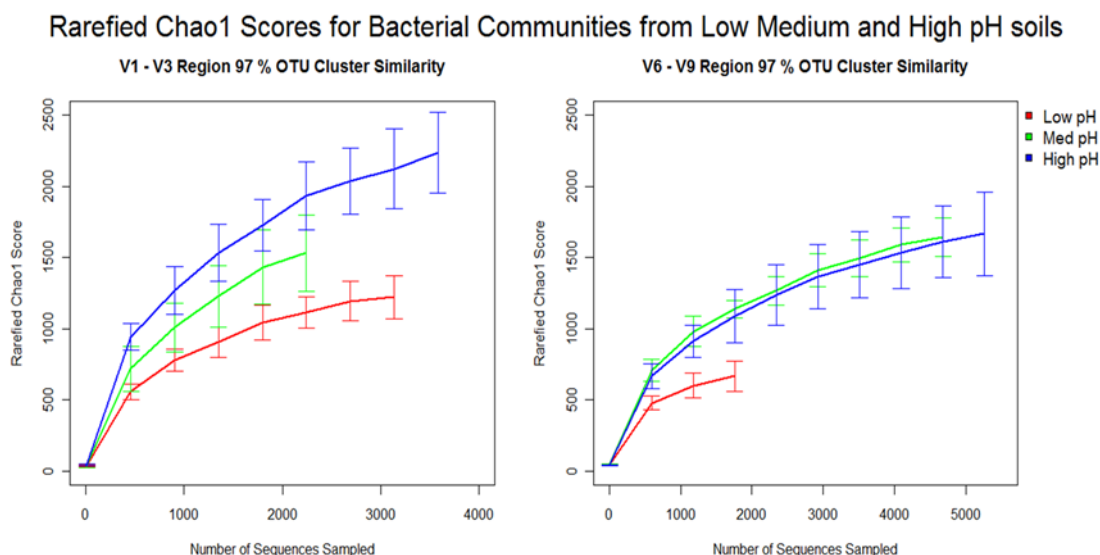
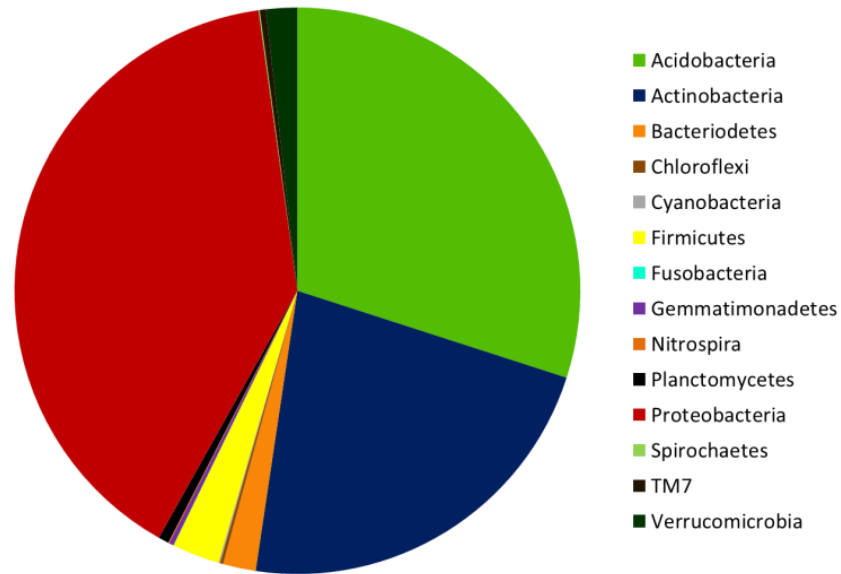


Figure 23: Mean Chao1 scores for rarefied subsamples of the two hypervariable regions targeted. The number of sequences sampled is limited to the smallest number of sequences within each pH class to avoid artificially inflating or deflating Chao1 scores for that group.

4.3.6 Differences in Community Composition between Variable Regions Targeted

Overall variation in reported community structure was compared between the variable regions targeted by pyrosequencing (Figure 24). Within the V6 – V9 sequence analysis acidobacterial OTUs made up 68.2 % of the total OTUs detected, compared with 29.9 % of the V1 – V3 region sequences. Proteobacteria and Actinobacteria OTUs were in higher proportional abundances within the V1 – V3 region analysis (39.6 % and 22.3 % respectively) compared with the V6 – V9 region analysis (15 % and 5 % respectively). No Chlamydiae OTUs were detected in the V1 – V3 region analysis but they made up 2 % of the total community in the V6 – V9 region analysis. Lower abundance taxa were similar in both analyses. However, taxa such as BRC1, an uncultured candidate division proposed by Derakshani *et al.* (2001), were not detected within the V1 – V3 region analysis but made up 0.02 % of the V6 – V9 region based community analysis. TM7 made up a small proportion of the total community when either region was analysed. TM7 is another uncultured candidate division characterised by environmental sequences only. It was originally isolated from a peat bog, from where the name was obtained (Torf Mittlere Schicht = peat, middle layer) (Rheims *et al.* 1996). Sequences identified as TM7 have been found in a wide range of environmental and clinical samples but little is known about their function or biogeography (Hugenholtz *et al.* 2001).

Proportional Abundance of Bacterial Phyla: V1 – V3 Region of the 16S rRNA Gene



Proportional Abundance of Bacterial Phyla: V6 – V9 Region of the 16S rRNA Gene

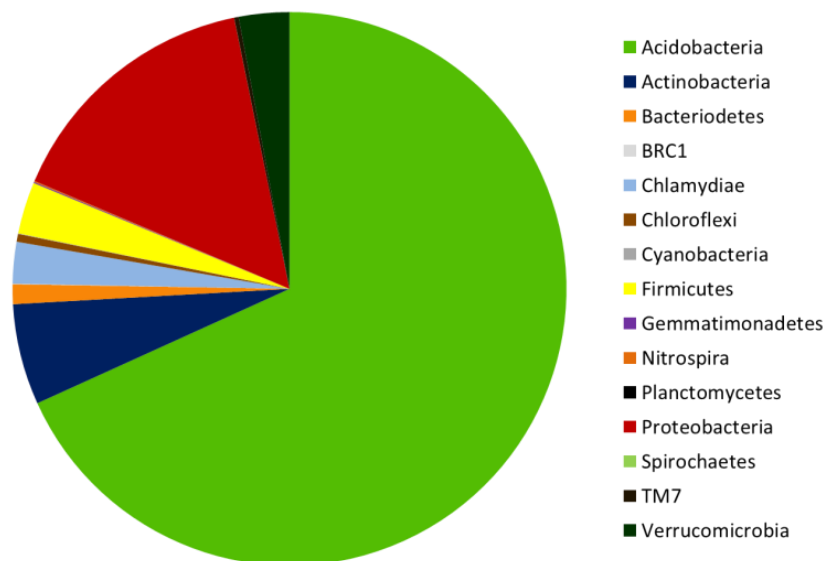


Figure 24: The proportional abundance of bacterial OTUs, at 97 % sequence similarity, with assigned taxonomies according to pyrosequencing of the A) V1 – V3 regions, or B) V6 – V9 regions, of the 16S rRNA gene.

4.3.7 Bacterial Community Composition Across a pH Gradient

To examine the changes in bacterial community composition across a pH gradient, proportional abundances were recalculated based upon the sample of origin (Figure 25).

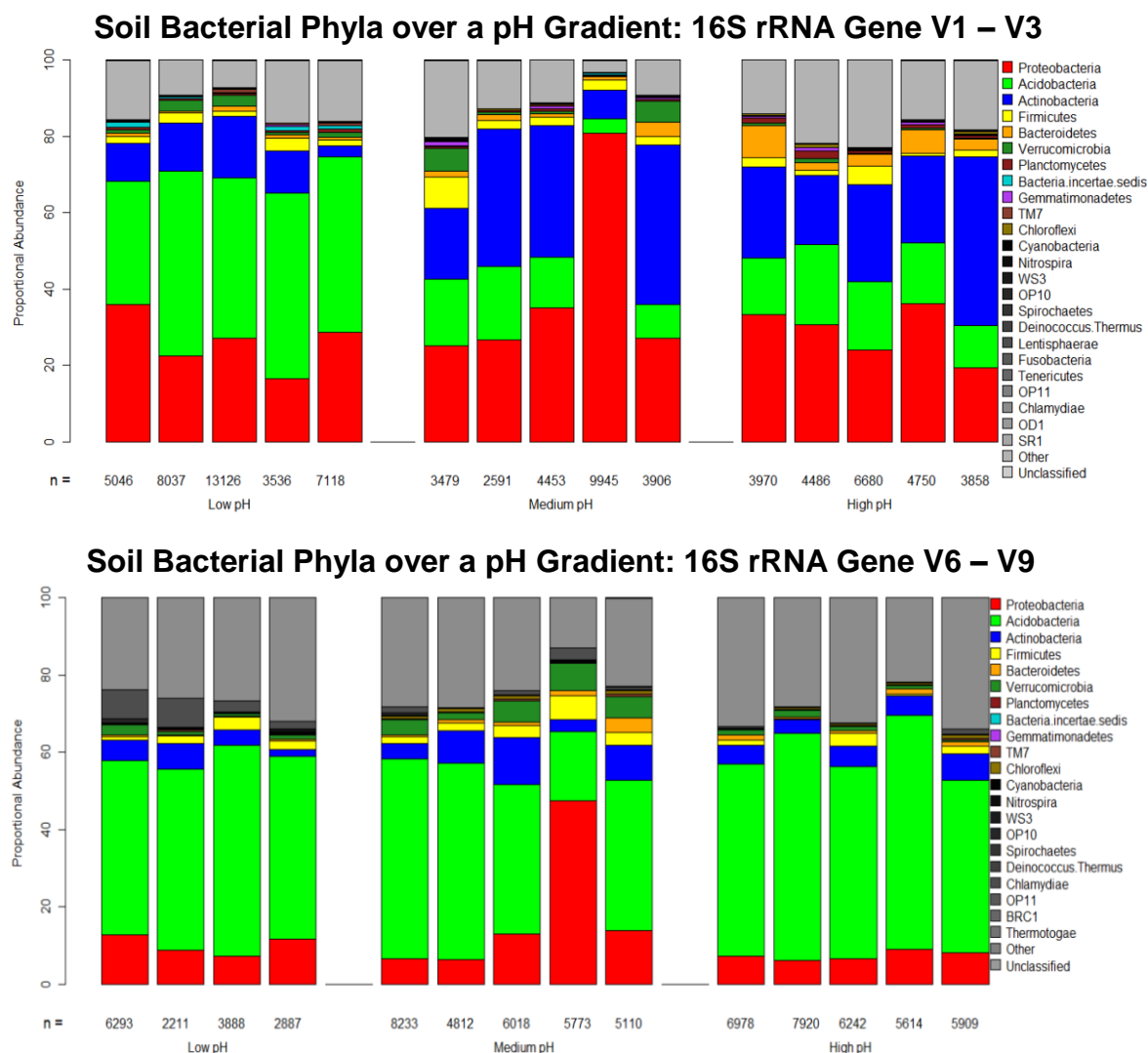


Figure 25: Proportional abundance of bacterial phyla in low, medium, and high pH soil samples. Colours were given to any phyla that represented > 1 % of the total community in any sample, and n denotes the number of quality checked sequences per sample. Unclassified OTUs were more prevalent within the V6 – V9 region analysis.

Within the V1 – V3 region analysis, Acidobacteria were in higher proportional abundances at low pH, and lower at medium and high pH. Little variation was observed in acidobacterial proportional abundance between the pH groups when the V6 – V9 region was analysed. The overarching trends in proportional abundance of acidobacterial subgroups and pH were similar when either set of variable regions was analysed (Appendix I). The majority of the acidobacterial subgroups were detected using both primer sets; however, a number of subgroups were detected at low abundances in the V1 - V3 region analysis but not the V6 - V9 region, these being groups 16 and 17. Groups 1, 2, 3 and 13 were detected using both primer sets and in each case they declined in proportional abundance with increasing pH. Groups 4, 5 and 6 showed a linear increase with pH. Groups 1 and 6 were the most abundant of the acidobacterial groups. Within the V6 – V9 region sequences, group 6 Acidobacteria were

shown to make up a considerably higher proportion of the total community than those detected using the V1 – V3 region sequences. This shows that even though overarching trends in proportional abundance of acidobacterial subgroups with pH were maintained irrespective of hypervariable regions targeted, the V1 – V3 region picked up more acidobacterial subgroups than the V6 – V9 region but the proportion of the total community made up by acidobacterial taxa was much greater in the V6 – V9 analysis. This was potentially due to the lower level of sequence dissimilarity associated with the V6 - V9 region (identified in Figure 20), thus resulting in a lower level of discrimination between closely related taxa.

Alpha and Betaproteobacteria were the most abundant of the Proteobacteria across both regions targeted (Appendix J), but the proportion of the total community attributed to the Alphaproteobacteria was substantially greater within the V1 – V3 region sequences than the V6 – V9 region sequences. This was true by several orders of magnitude and potentially explains the discrepancies in proportional abundance of Proteobacteria between the two 16S rRNA regions (Figure 25). Epsilonproteobacteria were detected in only two samples, at a low level of abundance using the V1 – V3 region primers, but not the V6 – V9 region primers. All classes of Proteobacteria (with the exception of Epsilonproteobacteria) were detected in each pH group. The proportional abundance of Alphaproteobacteria was higher in low and medium pH soils than high pH soils, while Betaproteobacteria were at higher proportional abundances in neutral soils than in those of an alkaline or acidic nature. The Deltaproteobacteria appeared to favour low and high pH soils according to data generated from the V1 – V3 region sequences, but when analysed using the V6 – V9 region sequences higher proportional abundances were observed in the medium pH samples. Gammaproteobacteria showed the same relationships with pH irrespective of the 16S rRNA gene targeted, that being an inverse unimodal distribution.

Actinobacteria were the third most abundant members of the bacterial communities studied. Large differences in proportional abundance were noted between the two primer sets used. In general terms, the V1 – V3 region primers detected a larger proportional abundance than the V6 – V9 region primers when the total community was taken into consideration. This was attributed to the

large proportional abundances of the Actinomycetales detected using the V1 – V3 region primers in comparison to the V6 – V9 region primers (Appendix K). However the distributions of proportional abundance of Actinomycetales across the pH classes were similar, being most abundant in more neutral soils. Using the V1 – V3 region as an indicator of taxonomic divergence, Acidimicrobiales appeared to have no distinct pattern with pH. However, when the V6 – V9 region was analysed few or no Acidimicrobiales OTUs were detected in the medium and low pH samples. All other phyla could not be confidently said to show distinct patterns with soil pH. It is worth noting that the V1 – V3 region primers exclusively detected the presence of sequences described as Rubrobacterales species in a single high pH sample and the V6 – V9 primers exclusively detected Coriobacteriales species in two medium pH samples (Appendix K). This is a further example of potential primer or region biases producing inconsistent results when different regions of the 16S rRNA gene are targeted for molecular based bacterial community profiling studies.

4.3.8 Analysis of Dominant Taxa Abundance Closer to the Species Level

Two phylogenetic trees were constructed to display changes in mean proportional abundance of dominant OTUs across the pH classes for each of the 16S rRNA gene regions analysed (Figure 26 and Figure 27). Full RDP classifications were added to terminal nodes, as were the mean proportional abundances of OTUs across each pH class. High resolution digital trees are supplied on the accompanying CD for ease of visualisation. Due to the higher sequence variability within the V1 – V3 region of the 16S rRNA gene highlighted previously, a larger number of OTUs were generated for these sequences. Both trees show that particular acidobacterial sub groups are prevalent in low pH samples, particularly groups 1 - 3, and that similar phylogenetic groupings were observed in both trees for both sets of hypervariable region sequences analysed. Alphaproteobacteria found at high abundances within low pH environments were in the orders Rhodospirillales (Family Acetobacteraceae) and Rhizobiales, while Betaproteobacteria, especially the order Burkholderiales, were most abundant within medium pH soils. Gammaproteobacterial taxa such as Xanthomonadaceae also exhibited this trend. In general actinobacterial lineages were evenly distributed across the medium and high pH soil types. However, notable OTUs that were predominantly detected in high pH soils were

Micromonosporaceae and Nocardioidaceae (Actinomycetales). A notable difference between the two trees was the clustering of phyla. Within the V1 – V3 tree, phyla tended to group well with each other, with the exception of the Solirubrobacterales, whereas within the V6 – V9 tree a number of OTUs clustered with phyla of a different taxonomic descent.

An indication of the number of OTUs unique to a pH class can be derived by viewing the bar charts attached to the terminal nodes of the tree plot. OTUs unique to the low pH samples were more common than those unique to the medium and high pH soils in both trees. Numerous OTUs were present in the medium and high pH soils but were absent in the low pH soils. Taxa showing this trend predominantly comprised Actinobacteria, and to a lesser degree, Proteobacteria lineages. This goes some way towards explaining the patterns seen in the UPGMA trees in Figure 22, as well as the higher Chao1 scores observed in the medium and high pH samples.

4.3.9 Shared OTUs Between and Within pH Groups

Appendix L i-iii and appendix M i-iii show patterns relating to shared or unique OTUs across and within pH groupings for the Acidobacteria (i), Alphaproteobacteria (ii) and Actinobacteria (iii) for V1 – V3 and V6 – V9 region sequences. Many of the OTUs identified as belonging to the Group 6 Acidobacteria were only found in the high and medium pH samples while the reverse was true for group 1 Acidobacteria, which were found predominantly in the low pH samples. Along with this, many of the OTUs identified as being unique to a particular pH class were detected in a majority of, but not all, samples of a particular pH class, suggesting that soils of similar physicochemical properties display different patterns of biodiversity when analysed closer to the species level. Again this is most notable with the group 1 and 6 Acidobacteria. Some OTUs identified as Alphaproteobacteria were shown to be shared predominantly across all samples, for example the Rhizobiales. However, others such as the Acetobacteraceae were predominantly found across most low pH samples, but not the medium and high pH samples. Actinobacterial lineages were shared between a majority of the medium and high pH samples, whereas Acidimicrobiales were predominantly restricted to the low pH soils. Slight discrepancies were noted when analysing the V6 – V9

region sequences. Acidobacterial group 1 was common to low and medium pH samples, as opposed to strictly low pH samples when V1 – V3 region sequences were analysed. However in both cases, acidobacterial group 1 OTUs were largely absent from high pH samples. The low number of Acidobacteria group 6 OTUs detected within the V6 – V9 region analysis highlights a lower level of sequence variability (as multiple sequences are grouped into a single OTU) and shows potential biases against this group as a result of primer or target region bias. As with the V1 – V3 based analysis, alphaproteobacterial OTUs were distributed across most medium and high pH samples, but a minority were found to be also present in the low pH samples. No alphaproteobacterial OTUs or groups were exclusive to a single pH class, but the majority were absent from the low pH samples, predominantly occurring within medium and high pH samples only. This was similarly true for the actinobacterial OTUs, with most being detected in at least two of the medium and high pH samples. Several OTUs were found across all pH groupings, but due to the low taxonomic level of classifications (i.e. to the phylum level) when using the V6 – V9 region sequences, OTUs found in all samples may have actually belonged to different groups of actinobacterial lineages closer to the species level.

Phylogenetic Tree Representing Dominant Bacterial Taxa and the Mean Proportional Abundance of OTUs in Low, Medium, and High pH Soils (V1—V3 Region)

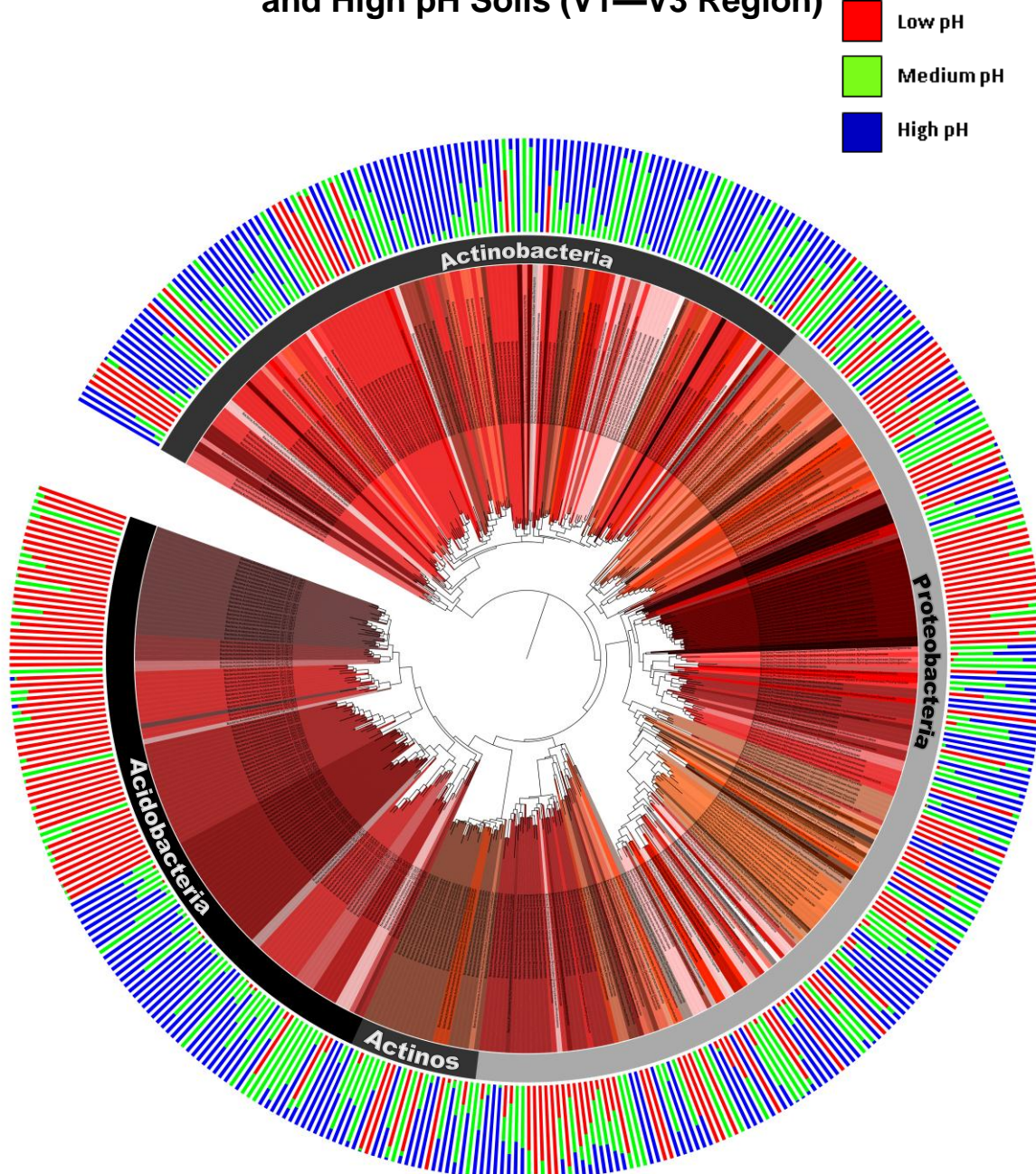


Figure 26: Phylogenetic placement of dominant OTUs derived from the V1 – V3 region of the 16S rRNA gene. The mean proportional abundance of each OTU within different pH groups is denoted by bar charts at terminal nodes. Clade colours represent groupings of classification close to the genus level. Coloured stripes represent Phylum level classifications. Generally, phyla clustered together in line with proposed phylogenetic placements of taxa suggesting tree based analysis using this region of the 16S rRNA gene is consistent with that of trees created with full length sequences (at least to the phyla level).

Phylogenetic Tree Representing Dominant Bacterial Taxa and the Mean Proportional Abundance of OTUs in Low, Medium, and High pH Soils (V6—V9 Region)

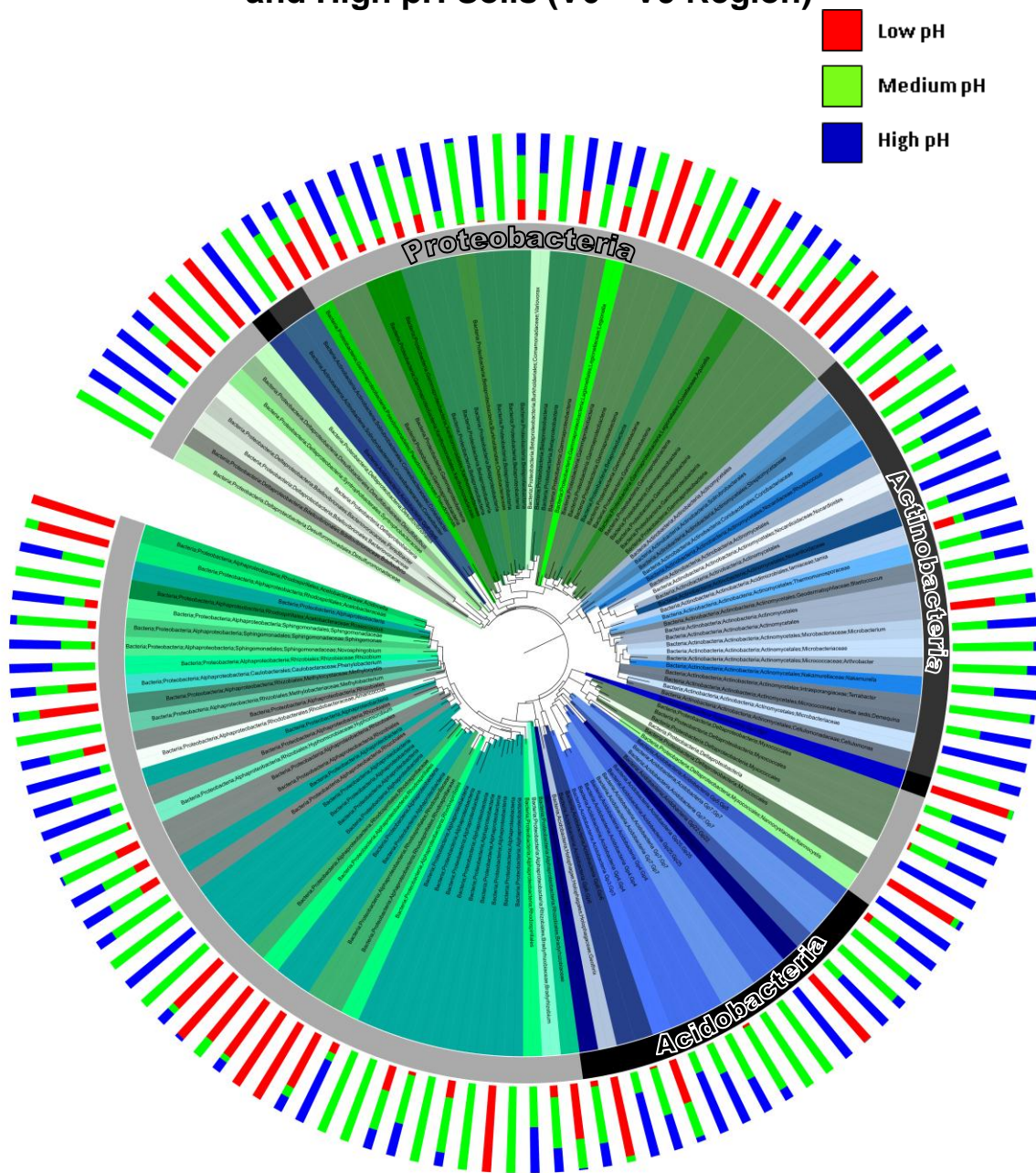


Figure 27: Phylogenetic placement of dominant OTUs with assigned taxonomies as derived from the V6 – V9 region of the 16S rRNA gene. Bar charts, representing the mean proportional abundance of that OTU across the pH groups and taxonomic classifications are displayed at the terminal nodes of the tree. Fewer OTUs were created during OTU clustering than with the V1 – V3 region tree, and assignments given to closely related taxa did not correspond to phylogenetic placement within this tree as they did in the V1 – V3 tree. Overarching patterns are similar however, with group 1 Acidobacteria being prevalent in low pH soils, and group 6 Acidobacteria common to high and medium pH samples.

4.3.10 Indicator Taxa of Soil pH Groups

Indicator “species” analysis was undertaken to identify the bacterial groups predominantly responsible for the changes in community composition across

the pH gradient. Indicator values are defined as a combination of relative abundance of taxa and their relative frequency of occurrence within the various groups of sites (Dufrene & Legendre 1996), in this case defined by soil pH. Significance values (p) are computed using a randomisation procedure as detailed in Dufrene & Legendre (1996). High indicator values for a specific lineage describe a high abundance in all samples of one pH group, and low abundances in samples of different pH groups. Values are expressed as a percentage. Tables 5 and 6 show indicator taxa for low, medium, and high pH samples as a result of indicator analysis of the V1 – V3 region and V6 – V9 region sequences. This was performed to highlight similarities and differences in overarching ecological trends described by the use of either variable region for pyrosequencing analysis.

Indicator Taxon	pH Group	Indicator Value	p
Acidobacteria (Group 13)	Low	1.00	0.001
Acidobacteria (Group 2)	Low	0.97	0.004
Acidobacteria (Group 1)	Low	0.95	0.005
incertae sedis (Ktedonobacteria)	Low	0.85	0.003
Acidobacteria (Group 3)	Low	0.82	0.002
Bacteroidetes (Bacteroidia)	Low	0.81	0.004
Verrucomicrobia (Opitutae)	Low	0.80	0.003
Acidobacteria (Holophaga)	Low	0.77	0.015
Acidobacteria (Group 15)	Low	0.75	0.022
Acidobacteria (Group 7)	Medium	0.85	0.002
Chloroflexi (Thermomicrobia)	High	0.88	0.001
Acidobacteria (Group 11)	High	0.87	0.004
Chloroflexi (Caldilineae)	High	0.84	0.005
Acidobacteria (Group 18)	High	0.80	0.015
Acidobacteria (Group 6)	High	0.74	0.003
Bacteroidetes (Sphingobacteria)	High	0.73	0.01
Nitrospira (Nitrospira)	High	0.72	0.009
Acidobacteria (Group 17)	High	0.68	0.009

Table 5: Bacterial indicator taxa (class level) for low, medium and high pH samples as determined by V1 – V3 pyrosequencing analysis. All indicator taxa had a p value less than or equal to 0.01.

Indicator Taxa	pH Group	Indicator Value	p
Acidobacteria (Group 13)	Low	1.00	0.003
Acidobacteria (Group 2)	Low	0.97	0.004
Acidobacteria (Group 1)	Low	0.95	0.003
Acidobacteria (Group 3)	Low	0.92	0.004
Acidobacteria (Group 7)	Medium	0.95	0.001
Proteobacteria (Betaproteobacteria)	Medium	0.93	0.009
Chloroflexi (Other)	Medium	0.80	0.004
Proteobacteria (Gammaproteobacteria)	High	1.00	0.003
Acidobacteria (Group 6)	High	0.51	0.002

Table 6: Bacterial indicator taxa (class level) for low, medium and high pH samples as determined by V6 – V9 pyrosequencing analysis. All indicator taxa had a p value of less than or equal to 0.01.

Acidobacterial taxa were predominantly responsible for the observed pH effect in low and high pH samples. Groups 13, 2, and 1 had high indicator values within the low pH samples according to analysis of either variable region set. These taxa had a high affinity for low pH soils and had low levels of occurrence in other pH groups. Group 6 Acidobacteria were the only acidobacterial lineage to be indicator taxa of high pH systems in both analyses. Betaproteobacteria were indicator taxa of medium pH systems within the V6 – V9 region analysis, but were not detected within the V1 – V3 region analysis. No taxa were shown to have contradictory patterns in terms of indicator species between the two analyses, but the V6 – V9 region analysis exhibited fewer taxa with significant indicator values. As the acidobacterial groups 1 and 6 were indicator taxa of the low and high pH samples respectively, and both were found in high abundances within samples of these natures (see Appendix I), it is hypothesised that these taxa are predominantly responsible for the pH effect observed within the ordinations displaying bacterial community dissimilarity (Figures 18-19).

4.3.11 Other Notable Rarer Taxa

Other phyla of considerably lower proportional abundances were also detected within this study, for example the Bacteroidetes, Firmicutes, and Planctomycetes. Appendix N (i – v) shows trees constructed for these phyla. A much larger proportion of the OTUs within these trees were endemic to a particular pH class than for the dominant phyla. This, however, is likely to be due to singletons arising as a result of the technical difficulty of sequencing the rare biosphere. One of the main differences between the two primer pairs used

when examining the rarer biosphere here is the fact that no Planctomycetes were found using the V6 – V9 primer pair. Coupled with this, other taxa such as the Bacteroidetes were seen in higher abundances in high pH soils than medium and low pH soils when the V1 – V3 region was used but were almost non-existent with the V6 – V9 region. The Verrucomicrobia were most abundant in neutral pH samples when analysed using the V6 – V9 region but no obvious pattern was seen when the V1 – V3 region was analysed. Because small scale pyrosequencing runs were undertaken to elucidate bacterial community structure in relation to soil pH changes, an in depth analysis of the variability in the rarer biosphere could not be confidently undertaken. Nevertheless, evidence suggests that pH may affect the community structure of the rarer biosphere, as well as dominant taxa.

4.4 Discussion

4.4.1 Overarching Patterns of Soil Bacterial Community Structure with Environmental Variance

Irrespective of the 16S rRNA region targeted, similar patterns in community dissimilarity were observed in relation to environmental variation. Furthermore, in concurrence with Chapter 3, similar environmental variables were found to be highly significant as community assembly determinants. Soil pH was shown to be one of the major predictors of bacterial community structure, as were the above ground plant community, loss on ignition, and soil moisture. This shows that, at taxonomic levels closer to that of the species, biogeographic patterns elucidated by amplicon based deep sequencing methods are similar to those achieved by molecular community profiling methods of lower taxonomic resolutions. Co-correlation between multiple environmental variables may again cause problems for determining the precise mechanisms by which soil bacterial communities are shaped within soils. However, the results of this chapter not only show that similar findings are observed when populations are examined at multiple levels of taxonomic resolution, but also provide a starting point for further research into these complex systems with the aim of unravelling potential co-correlations and determining the relationships between them and the patterns in soil bacterial biodiversity reported here.

Using phylogenetically based ordinations in reduced space, clear distinctions between the pH groups were observed, in concurrence with other studies (Fierer *et al.* 2007; Fierer & Jackson, 2006; Chu *et al.* 2010). Unifrac calculations are reliant upon robust tree building ability as distance matrices are calculated from clustered samples based upon the phylogenetic lineages shared between samples from one environment but not others (Lozupone & Knight, 2005). The ability to achieve an accurate phylogenetic placement of organisms using non-full length 16S rRNA sequences may potentially bias results. However, Unifrac based estimates of community dissimilarity using short sequencing reads are thought to be comparable to those generated from full length 16S rRNA reads (Liu *et al.* 2007).

When OTU sharing was examined using phylogenetic trees, acidobacterial OTUs were seen to be predominantly shared across most samples of one pH group, but absent from samples of others (e.g. low or high pH). This finding suggests that this group of organisms may not conform to Buffon's law of biogeography, that areas of similar environmental characteristics can, in some instances, exhibit different patterns of biodiversity. However, when analysed from a holistic point of view, relatively few OTUs from the Actinobacteria or Alphaproteobacteria were shared across all samples of one pH group and absent from another. This shows that, at higher taxonomic resolution than that provided by t-RFLP, biodiversity patterns of meta communities are not the same across all environmental gradients and thus it may not only be the "environment that selects" for particular bacterial taxa to proliferate. If this were the case, all samples of a defined environment would contain the same OTUs, but this was not observed. Thus it can be hypothesised that soil bacterial meta populations conform to Buffon's law of biogeography and the ecological laws governing their distribution in the environment are not substantially different from those for macro organisms when analysed at high taxonomic resolutions.

In terms of diversity (Chao1 richness), low pH soils were shown to harbour less diversity than medium and high pH soils, although the differentiation between diversity in medium and high pH soils was less distinct when the V6 – V9 region was used. This implies that patterns relating to bacterial diversity may be somewhat dependent upon the region of the 16S rRNA gene sequenced, and potentially explains the different shaped relationships observed in other

bacterial diversity studies (Fierer & Jackson, 2006; Chu *et al.* 2010). This also highlights the importance of a standardised approach to bacterial targeted pyrosequencing examinations of complex environmental communities.

Although overarching patterns in community dissimilarity and alpha diversity were similar between the two regions targeted, the actual community compositions generated differed considerably. The starkest contrasts between the two regions were in the proportional abundance of both the Acidobacteria and Actinobacteria. As the V1 – V3 region has a higher level of between sequence variability than the V6 – V9 region (Neefs *et al.* 1990), a larger number of OTUs was generated. The taxonomic assignments for these OTUs encompass a wider range of bacterial groups and subgroups, thus reducing the proportional abundance of the more dominant taxa.

4.4.2 Variable Region Choice Affects Reported Bacterial Community Structure

A study examining the effect of hypervariable region choice upon reported bacterial community composition showed that, using simulated short length bacterial reads sampled from human, mouse, and microbial mat samples, most regions of the 16S rRNA gene provide stable estimates of the phyla present in a sample, but further taxonomic assignment was highly sensitive to the region of the 16S rRNA gene sequenced (Liu *et al.* 2008). The authors suggest that the V6 region is unsuitable for providing accurate taxonomic assignments (as opposed to diversity indices) and that regions spanning the V2 – V3 region are optimum in this respect. Also, sequences spanning the V6 region are much less suitable for community clustering. Other studies have shown that incomplete 16S rRNA genes could not estimate OTU richness or define OTUs as reliably as full length sequences; however, regions towards the start of the 16S rRNA gene, such as V1 – V4, were optimum for use with bacterial community profiling studies by pyrosequencing (Kim *et al.* 2011). Despite this, other studies have found that the region of 16S rRNA gene sequenced did not significantly affect bacterial community assessments. The abundances of V3 and V6 region bacterial sequences, identified as belonging to the same organisms at varying taxonomic ranks (genus – phylum), showed a linear correlation to one another, i.e. identified taxa were found to be similar in abundance when either region

was used. Furthermore, similar patterns were found when either region was compared with full length sequences obtained from sequence repositories (Huse *et al.* 2008).

Here, the phylogenetically based placement of OTUs within the two trees representing the dominant biosphere was considerably different between the different 16S rRNA gene regions sequenced. The V1 – V3 region tree clustered organisms of the same phyla and closely related subgroups together, except for the placement of the actinobacterial lineage Solirubrobacterales. Lineages to which these organisms belong have been recently classified (Stackebrandt *et al.* 1997; Reddy *et al.* 2009) but are widespread in sequence collections from soil bacteria (Singleton *et al.* 2003). The low relatedness to other actinobacterial taxa reported by Singleton *et al.* (2003) suggests that *Solirubrobacter* and Solirubrobacterales are a distinct and novel genus and order (Singleton *et al.* 2003; Reddy *et al.* 2009). The divergence from other Actinobacteria within the tree reported here may have been magnified by the use of non - full length 16S rRNA gene sequences. Therefore, further work should include a more detailed investigation into the phylogenetic placement of these bacteria in relation to other soil bacterial phyla using full length sequence reads. OTUs classified using the V6 –V9 region sequences did not cluster within clearly defined phyla as well as those in the V1 – V3 region tree. Discrepancies in the tree may be attributed to the fact that taxa assignments are less robust using this region (Liu *et al.* 2008). In summary it is hypothesised that the V1 – V3 region sequences, although not full length, provide a more accurate portrayal of bacterial phylogenies and are in line with current placements of bacterial lineages based upon full length 16S rRNA gene sequence assessments. The V6 – V9 region provides a less taxonomically resolved, and thus less informative, portrayal of phylogenetic placement of taxa within this study.

The V1 – V3 and V6 – V9 regions are known to be the most divergent of all the 16S rRNA gene hypervariable sections, but the former has greater between sequence variability (Neefs *et al.* 1990) due to the faster rate of evolution in the V1 region (Schloss, 2010). This suggests a greater ability to discriminate between closely related taxa. As taxa assignment has become one of the primary goals for molecular community profiling studies (from which accurate diversity estimates can be calculated), pyrosequencing based estimates of

bacterial community composition should be performed using a standardised approach targeting the V1 – V3 region. This is a logical suggestion as pyrosequencing methods are unlikely to have the dramatic increase in read length production required for consistent full length 16S rRNA analysis in the near future (Huse *et al.* 2008). Coupled with this, computationally based taxon assignment methods such as the commonly used RDP classifier are based upon the use of sequences from previously identified organisms, and do not allow the identification of new taxa. De novo tree building methods aid in the classification of new taxa (Liu *et al.* 2008) and the large number of sequences produced per pyrosequencing run increases the probability of obtaining reads from novel taxa. However de novo tree building requires large computational expense to process the number of sequences that can be generated by pyrosequencing strategies. If these strategies are to be worth undertaking, any new taxa discovered by these methods should be included for use in the RDP classifier (or other methods reliant on an external database of sequences) to increase the accuracy of subsequent taxa assignments. A standardised sequencing approach is needed so that non-full length 16S rRNA sequences will encompass the same region once aligned. Prior to this, however, further analysis is required to evaluate the primer biases introduced when analysing complex bacterial populations targeting this region using pyrosequencing, so that a standard primer set can be used across studies.

4.4.3 Acidobacteria Show Strong Relationships with Soil pH

Acidobacterial groups 1 and 6 showed the greatest changes in proportional abundance with increasing pH and were deemed to be significant indicator taxa of the low and high pH samples respectively, independent of the variable regions targeted for analysis. Using either primer pair, the Group 1 Acidobacteria made up between 20 % and 45 % of the low pH communities. The group 6 Acidobacteria made up a large proportion of the V6 – V9 sequences, and were also a main constituent of the V1 – V3 region sequences (after Group 1). No group 1 Acidobacteria were detected within the high pH samples and very few were found within the medium pH samples. Similar findings have been reported in other studies, and suggest that the Acidobacteria are one of the most dominant groups within the soil biosphere (Janssen *et al.* 2006). Acidobacteria, as a phylum, has been shown to be negatively correlated

with available soil carbon content (Fierer, 2007) and in higher abundance within older soils than newly formed soils (Tarlera *et al.* 2008). A recent comprehensive survey of acidobacterial taxa across 88 soils of diverse composition using pyrosequencing shows similar results to this study as acidobacterial OTUs represented 30.9 % of all classified bacterial sequences (Jones *et al.* 2009). Furthermore, the proportional abundance of acidobacterial taxa in phylum level assessments decreased with increasing pH, as was seen in the V1 – V3 region analysis presented here, but other environmental variables (precipitation rates and carbon content) also had weaker but significant effects. Although both clone library analysis and pyrosequencing were carried out by Jones *et al.* (2009), the authors suggest that the primers used in near full length 16S rRNA clone library analysis of acidobacterial taxa underrepresent specific acidobacterial groups. Similar patterns to those found in this study were observed for subgroups, with 1, 2, 3, and 13 having negative correlations with pH and 6, 7, 11, and 18 having positive correlations. Jones *et al.* (2009) performed pyrosequencing using full length 16S rRNA amplicons, but mean read length was 232 bp, suggesting that variable regions 1 – 3 may have been used although this is not explicitly stated in the text. Here, different total abundances of acidobacterial taxa were observed when the V6 – V9 region was used, although similar patterns relating to acidobacterial subgroup abundance and pH were seen. This suggests that the total proportion of Acidobacteria detected within soils may be dependent upon the 16S rRNA region of choice, with the V6 – V9 region preferentially amplifying specific groups, most notably subgroup 6. Within this study, subgroup 6 made up, on average, 50 % of the total community of high pH samples when analysed using the V6 – V9 region, compared with 12 % using the V1 – V3 region (Appendix I).

4.4.4 Other members of the Dominant Biosphere

Other dominant members of the biosphere were not as strongly linked to pH as the acidobacterial lineages. However, Actinobacteria were found more often in medium and high pH soils, corroborating findings in Chapter 3. Actinobacteria are known to be indicative of good soil health (van Dijk, 2008) and a major constituent of grassland soils (Felske *et al.* 1997; Janssen, 2006). As in another pyrosequencing based study of soil bacterial biodiversity, Rubrobacterales were found in high pH soils (Lauber *et al.* 2009), although in this study a limited

number of sequences belonging to this group was detected. As sequence numbers per sample were much greater than those of Lauber *et al.* (2009) (between 3000 and 10000 in this study compared with 1501) and sequence reads were longer due to improved sequencing chemistry, misidentification in Lauber *et al.* (2009) or differences in primer bias in this study may explain these discrepancies.

In contrast to the study by Lauber *et al.* (2009), several environmental variables were identified as important determinants of bacterial community structure, rather than pH alone. This suggests that multiple environmental variables are important determinants of bacterial community structure, although pH was shown to be consistently the strongest explanatory variable for community dissimilarity. Again potential co-correlation must be taken into consideration. The direct mechanisms by which pH and co-correlates of soil pH affect bacterial community structure and diversity remain unclear, and this is an area for further research.

Of the Proteobacteria, the abundance of Alphaproteobacteria was shown to be high in low pH soils and decrease as soil pH increased. This is in contrast to the findings of Chapter 3, which provided the hypothesis that Alphaproteobacteria have a unimodal relationship with soil pH. Using the phylogenetic trees, OTUs designated as the Acetobacteraceae (Alphaproteobacteria) were shown to be good candidate taxa for the Alphaproteobacteria that were found in high abundances at low pH. It is possible that the larger depth of sequencing and higher taxonomic resolutions available from pyrosequencing provide a more detailed definition of taxa-environmental relationships closer to the species level, and that other factors may play important roles at these levels of taxonomic resolution. Even so, Betaproteobacteria were seen to be significant indicator taxa of the high pH samples, and Gammaproteobacteria were seen to be indicators of medium pH samples when the V6 – V9 region was analysed (but not the V1 – V3 region).

4.4.5 Conclusions

Similar ecological patterns were observed in relation to environmental variables and community dissimilarity to less taxonomically resolved studies of bacterial biogeography (Chapter 3) when either set of variable regions was analysed by

pyrosequencing. However, the community composition and relationship of individual taxa to soil pH were different between the two sets of hypervariable regions studied. This shows the need for further research into the optimum region(s) of the 16S rRNA gene to analyse via pyrosequencing, along with the minimum length needed to accurately portray bacterial community composition within complex environmental communities.

As OTUs detected in samples of a particular pH group were not shared across all samples of that pH group, soil bacterial populations are not the same across all samples with similar environmental characteristics when analysed at higher taxonomic resolutions. This highlights the advantage of using community analysis methods with higher taxonomic resolutions, such as pyrosequencing, over less taxonomically resolved methods such as t-RFLP when analysing populations from an ecological perspective. T-RFLP shows that dominant restriction fragments are ubiquitous (i.e. at the approximate phylum level, Chapter 3), whereas rarity is the norm when populations are analysed at higher levels of OTU stringency. This suggests that soil bacterial communities conform to Buffon's law of biogeography and are comparable to those of macro organisms in terms of community assembly.

The effect of soil bacterial biogeography has been assessed at multiple taxonomic resolutions, but other members of the soil micro biota may not adhere to the same patterns. Soil fungi are from another kingdom, are highly dominant within the soil ecosystem and fulfil many essential roles in degradation of organic matter and symbiotic links with plants. Similar analytical procedures are needed to understand how their communities are shaped in relation to the environment, and if they conform to similar or different ecological patterns to soil bacteria.

4.4.6 Further work

The Acidobacteria are undoubtedly an important constituent of soil bacterial populations. However, their strong relationship with soil pH potentially masks the effects of other environmental variables upon the total bacterial community. Large scale studies examining the effect of environmental variables upon bacterial community structure using pyrosequencing could remove sequences

identified as the Acidobacteria to examine the effects of environmental variability upon the remaining bacterial diversity.

Although this pyrosequencing method provides a large number of sequences per sample for the cost of the process, truly deep sequencing is sacrificed. We may obtain a representation of the rarer biosphere, but the truly rare biosphere will be missed. Moreover, although ecological patterns within the rarer biosphere were suggested here, the depth of sequencing achieved was not large enough to accurately assess these patterns. The importance of the rare biosphere within soil systems is under question as organisms which are extremely low in abundance may not greatly affect the overall function of a soil type, but there is some evidence to suggest this is a fallacy (Pester *et al.* 2010). Further work expressly targeting the rare biosphere, either by extremely deep sequencing or the artificial inflation of the rare biosphere, may provide useful insights into the biogeography of the rare soil bacterial biosphere and the resistance and resilience of soil function to environmental perturbation.

Chapter 5: Large Scale Fungal Community Analysis from Soils across the UK

5.1 Introduction

5.1.1 Background

Fungi are the most genetically diverse group of the Eukaryota (Buée *et al.* 2009) and are ubiquitous in the soil environment (Anderson & Cairney, 2004). They represent an essential functional component of terrestrial ecosystems, playing key roles in biogeochemical cycles as decomposers (Poll *et al.* 2010; Christensen *et al.* 1989) or through mutualistic relationships with other organisms (Chapela *et al.* 1994; Gianinazzi-Pearson, 1996). Additionally, numerous groups of soil fungi have been shown to harbour pathogenic properties important in relation to economically critical crops (Zellerhoff *et al.* 2010) and human infection (d'Enfert, 2009).

The understanding of soil fungal communities and how they operate under different environmental conditions is low compared to the wealth of knowledge acquired in relation to soil bacterial communities (Anderson & Cairney, 2004). In order to fully understand how ecosystems operate, a basic ecological framework is needed that encompasses all major soil based micro-organisms, not just the bacteria. Attempts to link above ground and below ground diversity or to gain insights into total soil function may be impossible or lead to erroneous conclusions until this has been achieved.

Elucidating the ecological gradients that help define the properties of soil fungal communities is a challenge due to the high taxonomic and ecological diversity within these systems (Buée *et al.* 2009). Sequence based assessments to enumerate the number of fungal OTUs per gram of soil vary depending upon the taxonomic marker gene of choice. However, both 18S rRNA gene and ITS based studies show that large numbers of OTUs are detected per gram of soil, with OTU numbers ranging between 600 (Fierer *et al.* 2007) and 3000 (Buée *et al.* 2009) at 97 % cluster similarity for the 18S rRNA gene and ITS region respectively. Within these studies, approximately 60 % of these OTUs identified by pyrosequencing are attributed to singletons, but the actual numbers of singleton sequences were relatively low. This implies that a small number of OTUs dominate samples, but the numbers of singleton sequences with a

greater than 3 % dissimilarity to other OTUs result in a high calculated diversity. Although Fierer *et al.* (2007) urge caution when comparing OTU richness between soil microorganisms, they suggest that OTU evenness is much lower in fungal populations when directly compared with bacterial populations: there is more variability in fungal communities than bacterial communities. Findings such as this have been attributed to the inconspicuous nature of fungi, and the complex array of dispersal mechanisms utilised by fungal organisms (Buée *et al.* 2009).

Small scale studies focusing upon specific groups of soil fungi (for example the ectomycorrhizal communities) have shown a strong link between soil fungal biogeography and plant biogeography, particularly within forest communities. This is, in part, due to the dependence of mycorrhizal fungi upon plant derived carbon (Azaizeh *et al.* 1995). Significantly reduced soil fungal diversity has been shown within rhizosphere associated soils compared to bulk soil (Smit *et al.* 1999), and a number of studies have demonstrated significant shifts in fungal community structure due to anthropogenic involvement (Mummey & Rilling, 2006; Tsui *et al.* 1998). This suggests that not only does a biogeographic effect exist within soil fungal communities, but also it is driven by surrounding environmental factors, some of which are known to affect the biogeography of plant communities (Partel, 2002). The land use history of a soil has been shown to be less important in terms of structuring fungal communities than the overarching edaphic conditions found there. For example, Lauber *et al.* (2008) showed, using a small scale, replicated study examining the effect of land use upon fungal and bacterial communities, that soil nutrient status had a larger impact upon fungal community structure than land use. Nevertheless, at the landscape scale it is likely that many edaphic conditions are, to some degree, correlated with land use history. Soil pH has been shown to affect community structure of a number of soil organisms, ranging from the well documented effect upon soil bacterial communities (Fierer & Jackson, 2006; Lauber *et al.* 2009; Chapter 3 - 4), through plant communities (Partel, 2002) and other soil microorganisms (Gubry-Rangin *et al.* 2011) but it is unclear whether fungal communities respond to environmental gradients in the same manner.

Few studies have explicitly examined the effect of soil pH upon fungal community structure in natural systems using a DNA based community profiling

method. Bacterial and archaeal based studies have observed a distinct pH effect, resulting in significant variation in community structure and diversity with soil pH (Gubry-Rangin *et al.* 2011; Lauber *et al.* 2009; Fierer & Jackson, 2006; Chapter 3- 4). This is potentially due to the effects of the intracellular stresses imposed upon organisms residing in environments at the extremes of pH which are outside the optimum for cellular function, for example enzyme denaturation, direct DNA damage (Cotter & Hill, 2003), disruption of transport proteins (Beales, 2004), increased solubility of toxic metals (Pina & Cervantes, 1996), and general inhibition of cellular functions. To overcome these restrictions, adaptations must have occurred to allow specific bacterial and archaeal populations to reside within these environments, resulting in some form of niche partitioning and an observable biogeographic effect (Fierer & Jackson, 2006). As soil fungi have been shown to be globally distributed and diverse (Fierer *et al.* 2007), some form of adaptation to pH stress can be expected also to have occurred within lineages belonging to this kingdom. If this is not the case, and no distinct pH effect is observed within fungal populations, then it must be hypothesised that soil fungal communities are structured in a distinctly different way to other dominant soil microorganisms.

Molecular methods used in studies examining soil bacterial communities are well documented (Marsh, 1999; Muyzer & Smalla, 1998; Ranjard *et al.* 2000). However, the same cannot be said for those targeting soil fungal communities. Many fungally targeted molecular based studies focus upon the identification of isolated organisms (Lahlali & Hijri, 2010; Bakri *et al.* 2010), or probing discrete communities (e.g. those in symbiotic relationships with roots) (Van Ooij, 2011). Phospholipid fatty acid analysis has been used to estimate fungal biomass in soils but is restricted by a limited number of fungal specific markers (Anderson & Cairney, 2004), and lacks taxonomic insight. T-RFLP analysis has recently become a popular choice for analysing soil fungal populations, utilising the 18S rRNA gene (Brodie *et al.* 2003; Mummey *et al.* 2005) or internally transcribed spacer (ITS) region (Lord *et al.* 2006; Mummey *et al.* 2005). The ITS region is often used due to its non-coding nature and thus faster rate of evolution (Anderson & Cairney, 2004), providing the ability to discriminate between closely related taxa (Lord *et al.* 2006). Furthermore the ITS region has been shown to have less bias towards particular fungal groups than the 18S rRNA

gene region, providing a more accurate representation of the fungal community under study (Anderson *et al.* 2003).

Few large scale studies have examined fungal biogeography at the landscape scale in an effort to define and quantify any taxa-environment relationship. As such, there is a need to highlight the environmental and edaphic variables that shape soil fungal populations in order to provide a basic ecological understanding of soil fungal populations, and soil microorganisms as a whole.

5.1.2 Aims

The principal aim of this study is to provide a large scale assessment of soil fungal biogeography at the landscape scale using an ITS targeted t-RFLP approach. Using 500 soil samples from two distinct and contrasting areas of soil and climatic conditions within the UK, relationships between environmental dissimilarity and fungal community dissimilarity were highlighted. Furthermore, small scale fungal clone library analysis of soil samples representing typical low, medium, and high pH soils within the UK was performed in order to give a taxonomic insight into the changes in fungal community composition over large ecological gradients.

5.2 Methods

5.2.1 Sample Selection

A stratified sampling design was used to select 500 soil samples from geographically and environmentally distinct areas of the UK. Fifty 1 km by 1 km areas, each encompassing a single habitat, were selected from the south of England and Wales, and a further 50 were selected from Scotland (Figure 28). Within each of these squares, 5 replicate soil cores (5 cm in diameter and 15 cm in depth) were obtained and processed under conditions outlined in Chapter 2. The rationale behind this sampling design was to maximise environmental dissimilarity between the two subsets from the north and south of the UK. To confirm a significant difference in environmental characteristics, Welch's two sample t-test was carried out. All environmental variables exhibited a significant difference between the two geographic areas ($p < 0.001$), except for annual cloud cover ($p = 0.5$) and soil phosphorus ($p = 0.09$). A Mantel test was used to show the correlation between distance and environmental dissimilarity (Mantel r

= 0.48, $p < 0.001$), and environmental dissimilarity was shown to be higher in the northern samples than in the southern samples. When environmental dissimilarity was correlated with increasing distance for both subsamples combined, total environmental dissimilarity was greater than that of either subset individually (Appendix O).

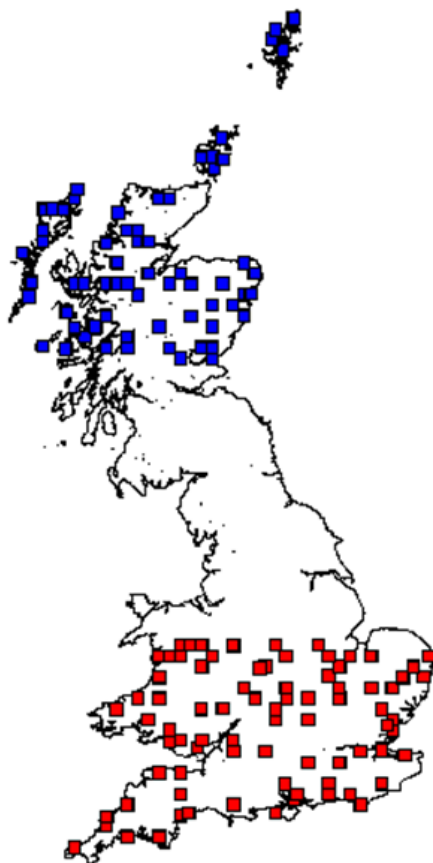


Figure 28: Geospatial locations of soil cores taken from the north of the UK (blue, $n = 250$) and the south of the UK (red, $n = 250$).

5.2.2 Nucleic Acids Extraction

Nucleic acid extraction was performed according to procedures outlined in section 2.2.2.

5.2.3 PCR Amplification of ITS Genes

Fungal t-RFLP profiles were generated using the method outlined in Lord *et al.* (2002). A ca. 650-700 bp region of DNA was amplified spanning the end of the 18S gene, the ITS1 region, the 5.8S gene, the ITS 2 region and the start of the 28S gene (Figure 29). The forward primer was EF3RCNL (5'-CAA ACT TGG TCA TTT AGA GGA -3') (Lord *et al.* 2002) labeled with fluorophore 6-FAM at

the 5' end, and ITS4 (5'-TCC TCC GCT TAT TGA TAT GC-3') (Lord et al. 2002) was used as the reverse primer.

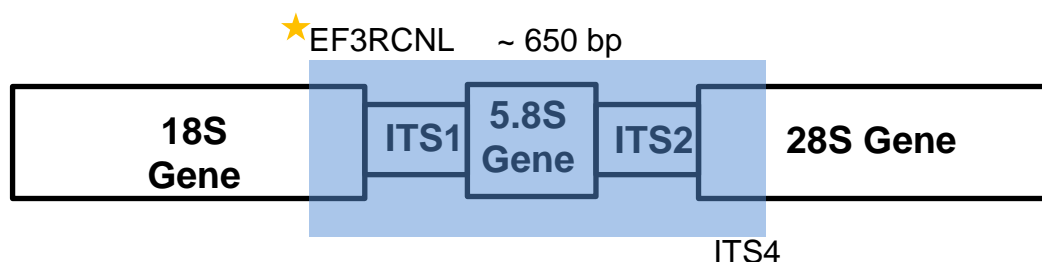


Figure 29: The ca. 650 -700 bp region amplified by primers EF3RCNI (6FAM) and ITS4.

PCR products were generated in 50 µl reactions using 5 U Taq polymerase (Sigma Aldrich, Dorset, USA), 8 mg ml⁻¹ bovine serum albumin (BSA) (New England BioLabs, Ipswich, USA), 0.2 µM of each primer (MWG Eurofins, Ebersberg, Germany), 0.6 µM dNTPs, 10 X TAQ polymerase buffer (Sigma-Aldrich, Dorset UK), and 2 µl of 1 in 10 diluted nucleic acids extract. The following PCR conditions were used. An initial denaturation stage of 94 °C for 5 minutes was followed by 30 cycles of 94 °C for 1 minute, 55 °C for 1 minute, 72 °C for 1 minute, and a final extension of 72 °C for 10 minutes.

PCR products were purified by gel purification using Sephadex G-50 (Sigma-Aldrich, Dorset, UK) in a 96 well format. Sephadex columns were hydrated with 300 µl molecular grade water and left to stand at 4 °C overnight. Sephadex plates were spun at 15,000 X g for 5 minutes to compact columns prior to reaction product purification at 15, 000 X g for 10 minutes. Agarose gels (1.2 %) were used to visualise successful amplification of all 500 samples.

5.2.4 T-RFLP

PCR product (50 ng) was digested in 10 µl reactions using 5 U of the restriction enzyme HaeIII (New England Biolabs, Ipswich, USA) at 36 °C for 4 hours. Complete digestion was verified by agarose gel electrophoresis. For t-RFLP analysis, 1 µl of digest was combined with 9 µl of denatured HI-DI formamide (Applied Biosystems, Paisley, UK) and 0.4 µl of Genescan-600 Liz size standard (Applied Biosystems, Paisley, UK). Analysis took place on an Applied Biosystems 3730xl capillary sequencer (Applied Biosystems, Paisley, UK). T-

RFLP profiles were analysed using the commercial software GeneMarker (Softgenetics, PA, USA). Binning of peaks was performed manually using an overlay of all profiles to reduce errors. Proportional abundance of each peak was calculated prior to statistical analysis.

5.2.5 Clone Library Analysis

Fungal targeted clone libraries were generated from 9 separate soils spanning a pH gradient. Three spatially independent soil cores from each pH class were selected from low pH ($n = 3$, pH 4.1 \pm 0.2), medium pH ($n = 3$, pH 6.02 \pm 0.06), and (high pH, $n = 3$ pH 8.12 \pm 0.11) (Appendix B). For each sample, three replicate PCR reactions were performed using primers EF3RCNL and ITS4, and then pooled. Pooled amplicons were gel purified using a QIAquick gel extraction kit (Qiagen, Crawley, UK) prior to ligation into a PCRII (dual promoter) vector (Invitrogen, Paisley, USA) according to the manufacturer's instructions. Plasmids were transformed into TOPO TOP10 chemically competent *E. coli* cells (Invitrogen, Paisley, USA). Positive clones were selected for growth on X-gal, IPTG, and 5X kanamycin selective media according to the manufacturer's instructions. Randomly selected white colonies were screened for the correct insert by colony PCR using primers M13f (5'- TGT AAA ACG ACG GCC AGT -3') (MWG Eurofins, Ebersburg, Germany) and M13r (5'- AGG AAA CAG CTA TGA CCA T -3') (MWG Eurofins, Ebersburg, Germany) under the following reaction conditions: 95 °C for 1 minute 30 seconds followed by 30 cycles of 94 °C for 30 seconds, 54 °C for 1 minute, 72 °C for 1 minute, and a final elongation stage of 72 °C for 6 minutes. Unidirectional sequencing reactions were carried out in 10 μ l reactions using 1 μ l of 1 in 10 diluted colony PCR product, 2 μ l Big Dye sequencing reaction mix (Applied Biosystems, Paisley, USA), 5 X Big Dye reaction buffer, 8 mg ml⁻¹ BSA and 0.2 μ M M13r, made up to 10 μ l with ddH₂O. Reaction conditions included an initial denaturing step of 96 °C followed by 35 cycles of 94 °C for 30 seconds, 55 °C for 20 seconds, and 60 °C for 4 minutes. Sequencing reactions were purified using Sephadex G-50 (Sigma-Aldrich, Dorset, UK) prior to analysis on an Applied Biosystems 3730xl capillary sequencer (Applied Biosystems, Paisley, USA).

5.2.6 Statistical analysis

Diversity indices were calculated for each t-RFLP profile using Simpson's 1 – D index according to the formulae outlined in Chapter 3. NMDS ordinations were carried out using the metaMDS function in the vegan package in R. Least squares regressions of scaled environmental variables (mean = 0, SD = 1) upon NMDS ordinations were carried out to define the proportion of variance in fungal community dissimilarity accounted for by each environmental variable individually using the envfit function in the ecodist package in R. Indicator taxa analysis was carried out using the indval function in the labDSV package after grouping t-RFLP profiles based upon the dominant plant species present at each sample site. Pairwise dissimilarity matrices used to analyse t-RFLP profile variability were carried out using a Bray-Curtis distance measure also using vegan in R. Pielou's evenness score was calculated using the following equation: -

$$J = \frac{H}{\log(S_i)}$$

Where J = Pielou's evenness (Pielou, 1966), H = Shannon's diversity index, and S_i = the number of TRFs within each sample.

5.2.7 Clone Library Sequence Analysis

Sequences were quality trimmed (mean quality score > 30) and vector clipped using pregap4 (Bonfield *et al.* 1994). Sequences > 100 bp in length after quality processing were clustered into OTUs at varying levels of similarity using a nearest neighbour approach in mothur (Schloss *et al.* 2009). Phylip (Felsenstien, 1993) formatted distance matrices were then created using a Jukes-Cantor method for use with diversity and OTU overlap studies. Sequences were identified using a standalone BLAST (Altschul *et al.* 1990) search against the UNITE database (Abarankov *et al.* 2010). Fully bootstrapped (n = 100) consensus based phylogenetic trees were created using the phylip cluster of programs, in particular seqboot, dnaml, and consense. Phylogenetic analysis was performed in the biolinux 6 environment (Field *et al.* 2006).

5.3 Results

5.3.1 Environmental Effects upon Fungal Diversity

Simpson's diversity ($1 - D$) was calculated for each t-RFLP profile and correlated against each measured environmental variable to highlight any relationships between fungal diversity and the environment (Appendix P). No measured environmental variable had a significant correlation with fungal diversity, suggesting that at the landscape scale, fungal diversity is independent of environmental variation. Diversity scores for each sample were consistently high, suggesting that samples comprised numerous TRFs with a high degree of between sample variability and there were few dominant TRFs across all samples.

Although alpha diversity showed no significant relationship with any environmental variable, beta diversity was examined in relation to increasing pH as was performed in Chapter 3 to compare fungal and bacterial community responses to changing pH (Figure 30). Again as pH increased, beta diversity decreased suggesting a higher degree of community variability in low pH systems than in high pH systems.

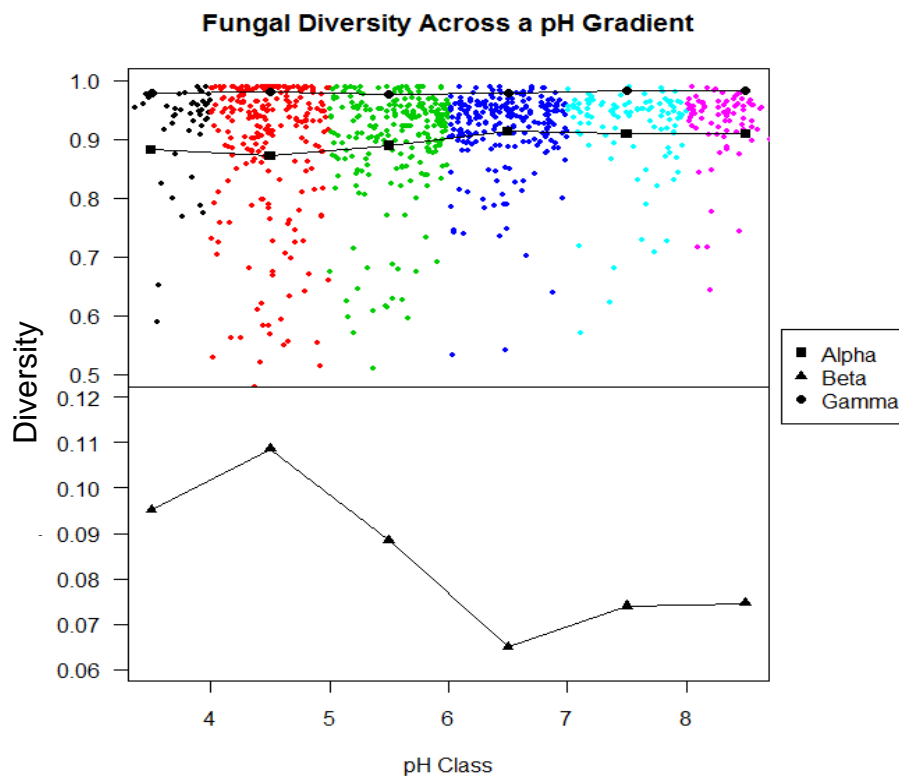


Figure 30: Scatter plot showing that alpha (Simpson's $1 - D$) and gamma diversity had no relationship with increasing pH, but beta diversity decreased as soil pH increased suggesting that fungal communities were more variable within low pH soils than high pH soils (Linear $R^2 = 0.47$, $p < 0.001$).

5.3.2 Environmental Effects upon Fungal Community Structure

A NMDS ordination was constructed from all t-RFLP profiles to which environmental variables were fitted in the form of a least squares regression (Figure 31). The relative effect of each environmental variable upon fungal community dissimilarity is shown in Table 7. No numerical environmental variables explained more than 33 % of the variance in the fungal community data matrix, and most variables returned R^2 values of less than 0.1. Therefore no strong significant links could be made with conviction between environmental variability and fungal community dissimilarity. The dominant plant species explained the largest proportion of the variance within the plot ($R^2 = 0.33$, $p < 0.001$). Again, correlation amongst predictor values must be taken into consideration when interpreting results, but this plot indicates that the distribution and community structure of soil fungi are not totally random and are in some way influenced by the environment.

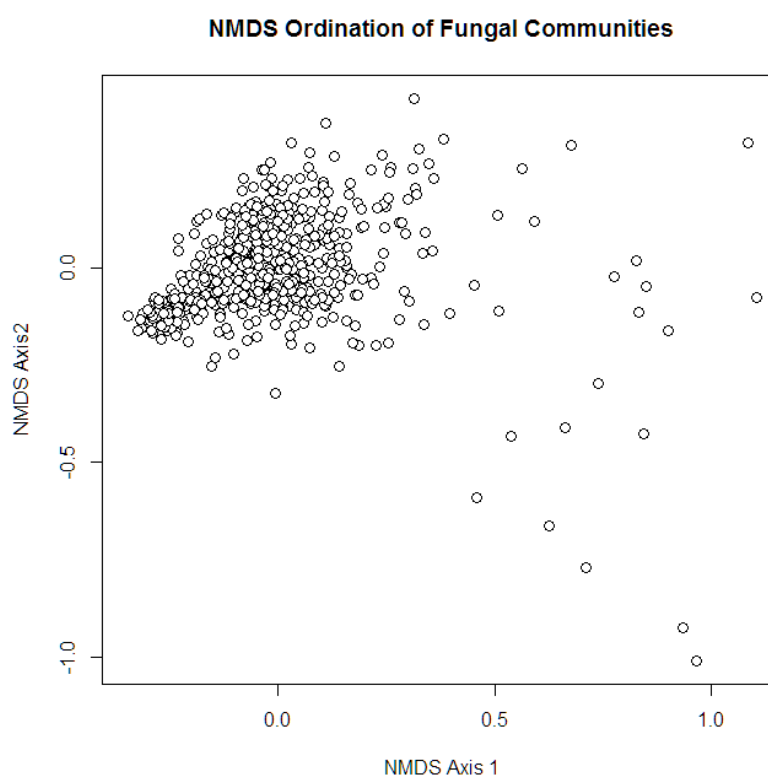


Figure 31: NMDS ordination of fungal community dissimilarity generated from 500 t-RFLP profiles from the north and south of the UK. Environmental variables were fitted to this plot in the form of a least squares regression approach.

Environmental Variable	r ²	p
Dominant Plant Species	0.3321	0.001
Broad Habitat Description	0.1269	0.002
Aggregate Vegetation Classification	0.0758	0.001
Annual Sunshine (Hrs)	0.055	0.001
Soil Texture Analysis	0.0549	0.001
Annual Temperature (°C)	0.0534	0.001
Loss on Ignition (g)	0.0377	0.001
Plant DCA Axis 1 Scores	0.0305	0.001
Plant DCA Axis 2 Scores	0.0282	0.002
Soil pH	0.0244	0.003
Soil Moisture (%)	0.0226	0.006
Plant DCA Axis 3 Scores	0.0194	0.009
Altitude (m)	0.017	0.012
Carbon Content (%)	0.0121	0.039
Phosphorus Content (%)	0.011	0.065
Annual Rainfall (mm)	0.0102	0.063
Annual Cloud Cover (Hrs)	0.0096	0.085
Carbon:Nitrogen	0.0048	0.286
Nitrogen Content (%)	0.0017	0.627

Table 7: Effect of environmental variables upon fungal community dissimilarity as defined by a least squares regression upon the NMDS ordination in Figure 31. Variables are ordered in terms of importance (denoted by decreasing R² value). Most variables had weak but significant effects upon fungal community dissimilarity, with the dominant plant species being the best predictor.

To determine if any TRFs had strong links with particular plant species, indicator taxa analysis was undertaken. Indicator values (constrained between 0 and 1) show how restricted each TRF is to a particular habitat, in this case defined by the dominant plant species. Indicator values are defined as a combination of relative abundance of taxa and their relative frequency of occurrence within the various groups of sites (Dufrene & Legendre 1996), in this case defined by soil pH. Significance values (p) are computed using a randomisation procedure as detailed in Dufrene & Legendre (1996). Values are expressed as a percentage with high indicator values indicating that the TRF in question is found almost wholly within all samples of one type, and is largely absent from samples of another grouping (Table 8).

A grouping variable was assigned to each t-RFLP profile to represent the dominant plant species present at the sample site. Table 7 summarises the 24 TRFs that had a significant link with the dominant plant species present at each sample site. Many of the plant species found to have significant links with fungal

TRFs are predominantly found in moist soils, with the majority having acidic properties. Some, such as birch and bracken, are found in acidic but well drained soils.

TRF (bp)	Dominant Plant Species	Typical Plant soil conditions	Indicator Value	P value
75	<i>Betula pendula</i>	Acidic, well drained	0.0719	0.039
163	<i>Calluna vulgaris</i>	Acidic Heath	0.058	0.006
216	<i>Cynosurus cristatus</i>	Neutral pH	0.0748	0.001
225	<i>Cynosurus cristatus</i>		0.0401	0.036
499	<i>Epilobium sp</i>	Acidic (Some sp)	0.1339	0.003
210	<i>Erica tetralix</i>	Acidic, Damp	0.0409	0.001
148	<i>Eriophorum angustifolium</i>	Acidic, Damp	0.0633	0.002
213	<i>Eriophorum angustifolium</i>		0.0466	0.007
178	<i>Eriophorum angustifolium</i>		0.046	0.002
248	<i>Festuca arundinacea</i>	Slight Acidic, Damp	0.0474	0.028
254	<i>Festuca arundinacea</i>		0.0427	0.021
402	<i>Festuca arundinacea</i>		0.0361	0.029
175	<i>Juncus articulatus</i>	Alkaline, Damp	0.0927	0.007
360	<i>Juncus effusus</i>	Acidic, Damp	0.1563	0.027
222	<i>Juncus effusus</i>		0.0496	0.002
401	<i>Juncus squarrosus</i>	Acidic, Heath	0.0462	0.022
71	<i>Lolium multiflorum</i>	Neutral, Damp	0.0975	0.003
397	<i>Lolium perenne</i>	Neutral, Damp	0.0829	0.001
398	<i>Lolium perenne</i>		0.0581	0.011
403	<i>Lolium perenne</i>		0.0473	0.002
391	<i>Nardus stricta</i>	Acidic, Damp, Organic	0.0909	0.028
234	<i>Polygonum aviculare</i>	Damp, but widespread	0.032	0.035
122	<i>Pteridium aquilinum</i>	Low – Neutral pH, Drained	0.1624	0.009
115	<i>Schoenus nigricans</i>	Damp, Alkaline	0.1992	0.002

Table 8: Indicator species analysis defining a statistical link between specific TRFs and the dominant plant species present at a sample site. Indicator values are constrained between 0 and 1 with 1 representing TRFs only found in all samples of one topography (i.e. plant species).

Although certain TRFs had significant links with particular plant species, the indicator values for each were comparatively low. This shows that although specific TRFs were found in higher abundances within soils harbouring specific dominant plant species, these TRFs were also found at similar proportional abundances in other soils but less frequently.

First and second NMDS axis scores for the TRF vectors were plotted to determine the identity and quantity of TRFs associated with positive and negative NMDS axis sample scores. Appendix Q shows that relatively few TRFs were associated with samples exhibiting positive first axis NMDS scores, these being 17 TRFs below 200 bp in length, and 8 that were greater than 200 bp. A

large number of the remaining TRFs were strongly associated with negative NMDS first axis scores. No obvious links between TRFs identified in Table 7 and the first or second axis vector scores were seen. Therefore, although dominant plant species was highlighted as important in terms of structuring fungal communities, specific patterns relating to plant communities at this spatial scale could not be identified.

A large number of TRFs were associated with negative first axis NMDS ordination scores, and relatively few with positive first axis ordination scores (Appendix Q). This suggests that samples with negative first axis NMDS scores are more diverse than those with positive NMDS scores. To further investigate this finding, diversity and evenness were plotted for each sample according to the first axis NMDS score (Figure 32).

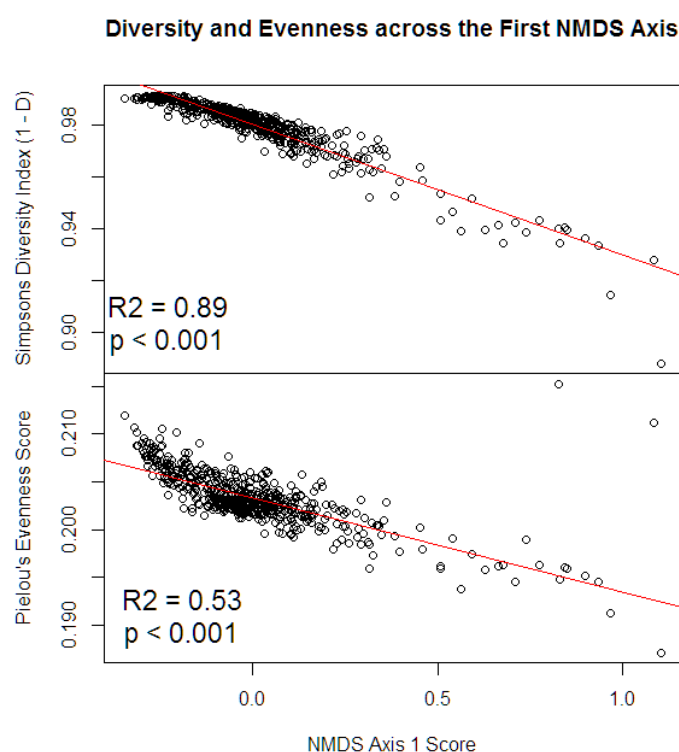


Figure 32: Simpson's diversity index and Pielou's evenness score decrease significantly when moving from the left to the right along the first axis of the NMDS ordination in Figure 31.

Each sample produced a high diversity score, with the majority falling between 0.94 and 0.98. A decrease in alpha diversity could be seen when moving along the first axis of the NMDS ordination. Negative first axis scores were associated with high diversities and higher evenness scores. In conjunction with Appendix Q, this shows that samples to the left of the ordination comprised communities

with large numbers of TRFs, few of which were dominant. Those to the right of the ordination were slightly less diverse but had lower evenness scores, suggesting that fewer, more dominant TRFs populated these communities.

Although there was a high level of fungal community dissimilarity, none of the environmental variables could explain a large proportion of this dissimilarity at the landscape scale. The samples from the north of the UK were then analysed separately from the southern samples in order to assess fungal community structure across shorter environmental gradients. Table 9 shows the relative effects of environmental variables upon fungal community dissimilarity as the result of a least squares regression fit to NMDS ordinations calculated from northern and southern t-RFLP profiles independently.

Northern Samples			Southern Samples		
Environmental Variable	R ²	p	Environmental Variable	R ²	p
Plant Species	0.3149	0.005	Plant Species	0.3984	0.001
Loss on Ignition	0.2181	0.001	Broad Habitat Description	0.1606	0.002
Broad Habitat Description	0.2159	0.001	AVC	0.1108	0.002
Nitrogen:Carbon	0.1583	0.001	Soil pH	0.1008	0.001
Soil Moisture (%)	0.1419	0.001	Annual Rainfall (mm yr ⁻¹)	0.0785	0.001
AVC	0.1392	0.001	Phosphorus Content (%)	0.0644	0.001
Carbon Content (%)	0.1133	0.001	Plant DCA Axis 1 Scores	0.0525	0.001
Soil Texture Analysis	0.1125	0.001	Loss on Ignition	0.0509	0.002
Soil pH	0.1062	0.001	Altitude (m)	0.0504	0.001
Nitrogen Content (%)	0.0895	0.001	Carbon Content (%)	0.0452	0.003
Annual Rainfall (mm yr ⁻¹)	0.0891	0.001	Nitrogen Content (%)	0.0436	0.003
Annual Sunshine (Hrs yr ⁻¹)	0.0783	0.001	Soil Texture Analysis	0.0424	0.021
Cloud Cover (Hrs yr ⁻¹)	0.0361	0.019	Soil Moisture (%)	0.0414	0.002
Annual Temperature (°C)	0.0299	0.041	Nitrogen:Carbon	0.0214	0.046
Plant DCA Axis 2 Scores	0.0295	0.033	Plant DCA Axis 3 Scores	0.0179	0.069
Plant DCA Axis 3 Scores	0.0235	0.063	Carbon:Nitrogen	0.0166	0.092
Carbon:Nitrogen	0.0131	0.237	Annual Temperature (°C)	0.0162	0.097
Altitude (m)	0.011	0.285	Annual Sunshine (Hrs yr ⁻¹)	0.0125	0.165
Phosphorus Content (%)	0.0039	0.603	Plant DCA Axis 2 Scores	0.0076	0.313
Plant DCA Axis 1 Scores	0.0028	0.754	Cloud Cover (Hrs yr ⁻¹)	0.0011	0.76

Table 9: The relative importance of environmental variables in determining fungal community structure over shorter environmental gradients determined by analysing the northern and southern samples independently.

Dominant plant species again had the highest R² value for both subsets of samples. Within the northern samples, loss on ignition, broad habitat description, N:C, soil moisture, AVC, soil carbon content, soil texture, and soil pH each explained > 10 % of the variation within the plot. Within the southern

samples broad habitat description, AVC, and soil pH each explained > 10 % of the variation in the plot.

Due to the high diversities and small environmental effects upon fungal community dissimilarity, the total inter t-RFLP profile variability was investigated to determine whether this could be a confounding factor. A pairwise dissimilarity matrix was calculated for all fungal t-RFLP profiles, as was a second for bacterial t-RFLP profiles generated in Chapter 3 from the same soil samples. Both the mean pairwise dissimilarity and total inter profile differences were greater for the ITS targeted t-RFLP profiles than the bacterial t-RFLP profiles (Figure 33).

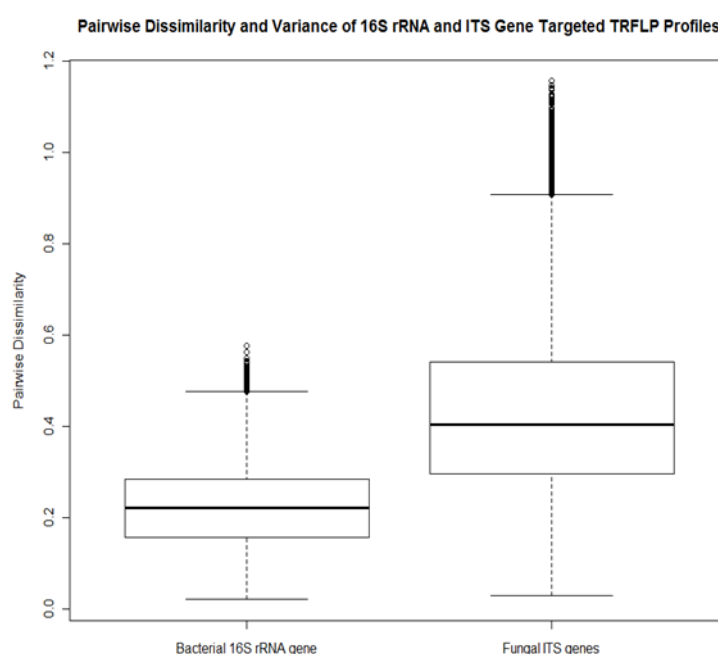


Figure 33: Boxplots showing the mean bacterial and fungal t-RFLP profile dissimilarity for the same 500 soil samples. Total fungal t-RFLP profile dissimilarity was significantly larger than bacterial profile dissimilarity (t-test $p < 0.005$). Bold black lines represent the median of each data set, while boxes represent the inter quartile range (IQR). Error bars denote the range of the data after the removal of statistical outliers while points represent statistical outliers as defined by > 3 times the IQR.

Coupled with the knowledge that the ITS gene is non-coding and thus more variable than the 16S rRNA gene, it could be inferred that this t-RFLP method potentially discriminates at greater taxonomic resolutions than the 16S rRNA gene t-RFLP approach, or that fungal community dissimilarity is far greater than that of soil bacterial populations.

The effect of environmental variables upon fungal community dissimilarity was small in comparison with findings presented in earlier chapters describing the

relationship between the environment and bacterial community structure. When fungal communities were analysed over shorter environmental gradients, the relative importance of several environmental variables increased. This suggests that fungal community structure is not wholly independent of the environment; however, the relationship is minimal when examined at the landscape scale with this t-RFLP method. A Mantel test was performed to correlate t-RFLP profile dissimilarity with environmental dissimilarity to determine whether a statistical link could be found. A significant positive Mantel R statistic was produced (Mantel $R^2 = 0.13$, $p < 0.001$), again suggesting small but significant effects of environmental dissimilarity upon fungal community dissimilarity.

As this community profiling method is not taxonomically informative and has an uncertain level of taxonomic resolution, clone library analysis was undertaken on geographically independent soils representing three environments with divergent physicochemical properties.

5.3.3 Clone Library Analysis

Soil samples representative of a range of different land use types and spanning a pH gradient were used to analyse fungal communities. Three low pH samples (pH 4.1 \pm 0.2), 3 medium pH samples (pH 6.02 \pm 0.06) and 3 high pH samples (pH 8.12 \pm 0.11) were examined using cloning and sequencing. One high pH sample did not produce viable clone libraries after multiple attempts and was thus abandoned from subsequent analysis. The soil samples selected represented 3 environmentally distinct habitats largely described by differences in soil pH. Low pH samples were typical organic acidic soils, while high pH soils were predominantly from improved grasslands or agricultural land. Medium pH samples were a mixture of improved and organic soils but had intermediate pH values. Sequences from each of the clone libraries were clustered into OTUs at increasing levels of similarity (from 60 – 100 % at 1 % increments) and Simpson's diversity indices ($1 - D$) calculated for a total of 748 sequences. Fungal diversity increased with increasing similarity level for OTU clustering for all pH groups. However, generally, fungal communities were more diverse in high and medium pH soils than low pH soils (Figure 34).

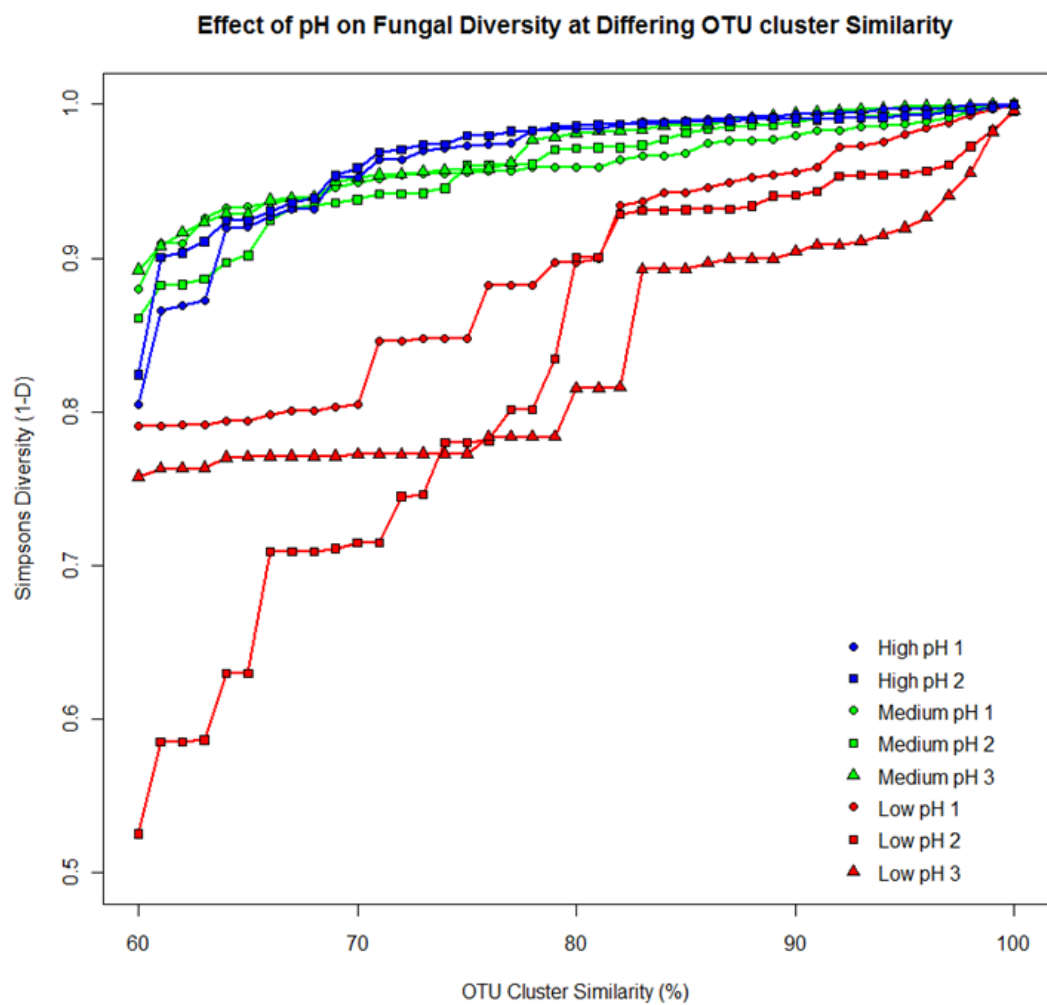


Figure 34: Simpson's diversity index for fungal clone libraries calculated at OTU clustering levels from 60 % – 100 % at 1 % increments.

Principal component analysis was undertaken from distance matrices of all aligned samples to show community dissimilarity at differing levels of OTU similarity (Figure 35).

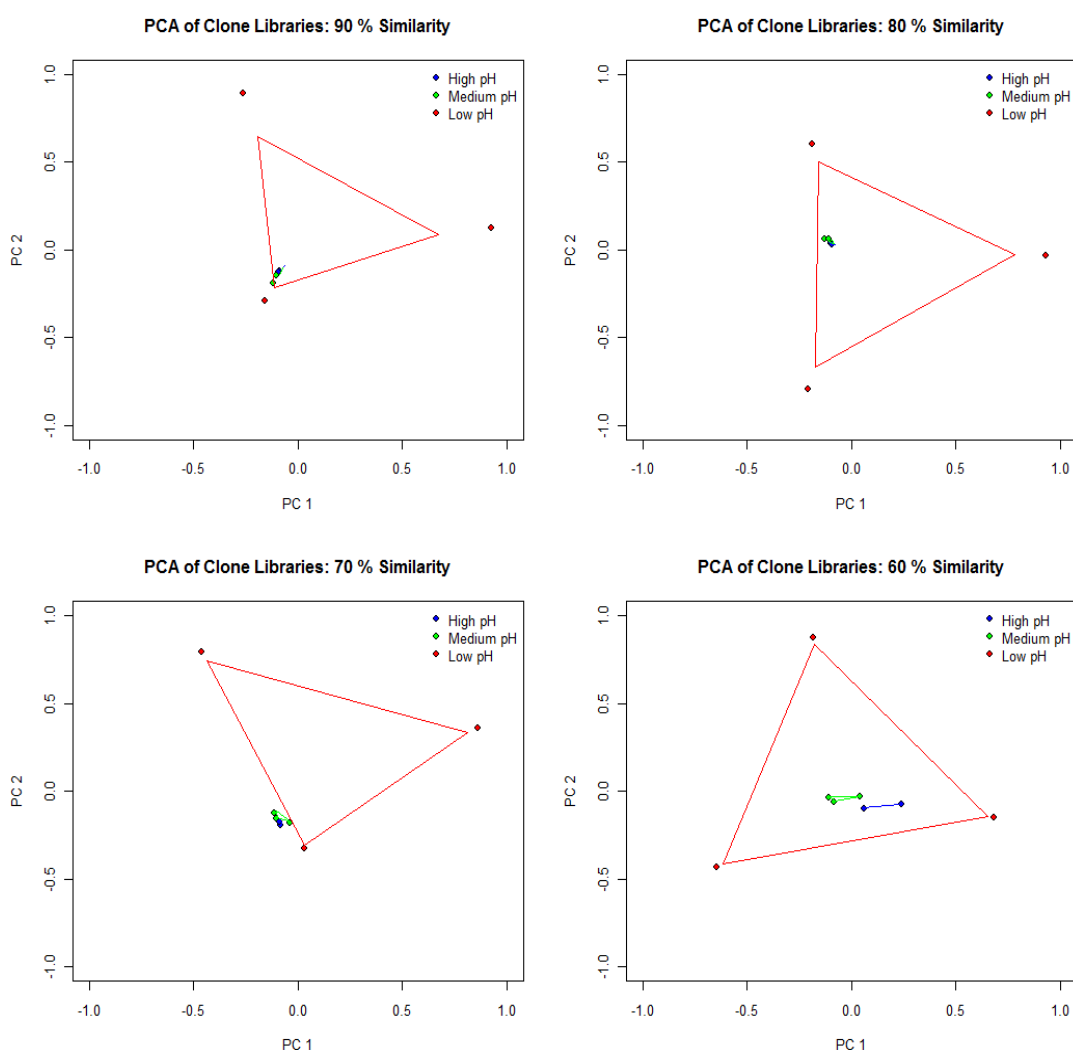


Figure 35: Principal component analysis showing the overall community dissimilarity of each clone library at decreasing levels of OTU similarity.

At all levels of OTU clustering, the medium and high pH samples were more similar in community structure than the low pH samples. Low pH samples were separated from one another by large distances across both principal components, implying a wider variation in community structure among samples from this group. Figure 34 and Figure 35 show that fungal communities found in medium and high pH soils are more diverse but less variable than those in the low pH samples. To expand upon this finding, Venn diagrams were constructed to display the OTU overlap between pH groups at different levels of OTU clustering (Figure 36). Medium and high pH samples consistently shared the most OTUs at all levels of clustering, and low pH samples generally had fewer

unique OTUs than the medium and high samples. At the highest levels of clustering (> 90 %) very few taxa were shared between any groups, suggesting a large variation in community structure between samples when analysed closer to the species level.

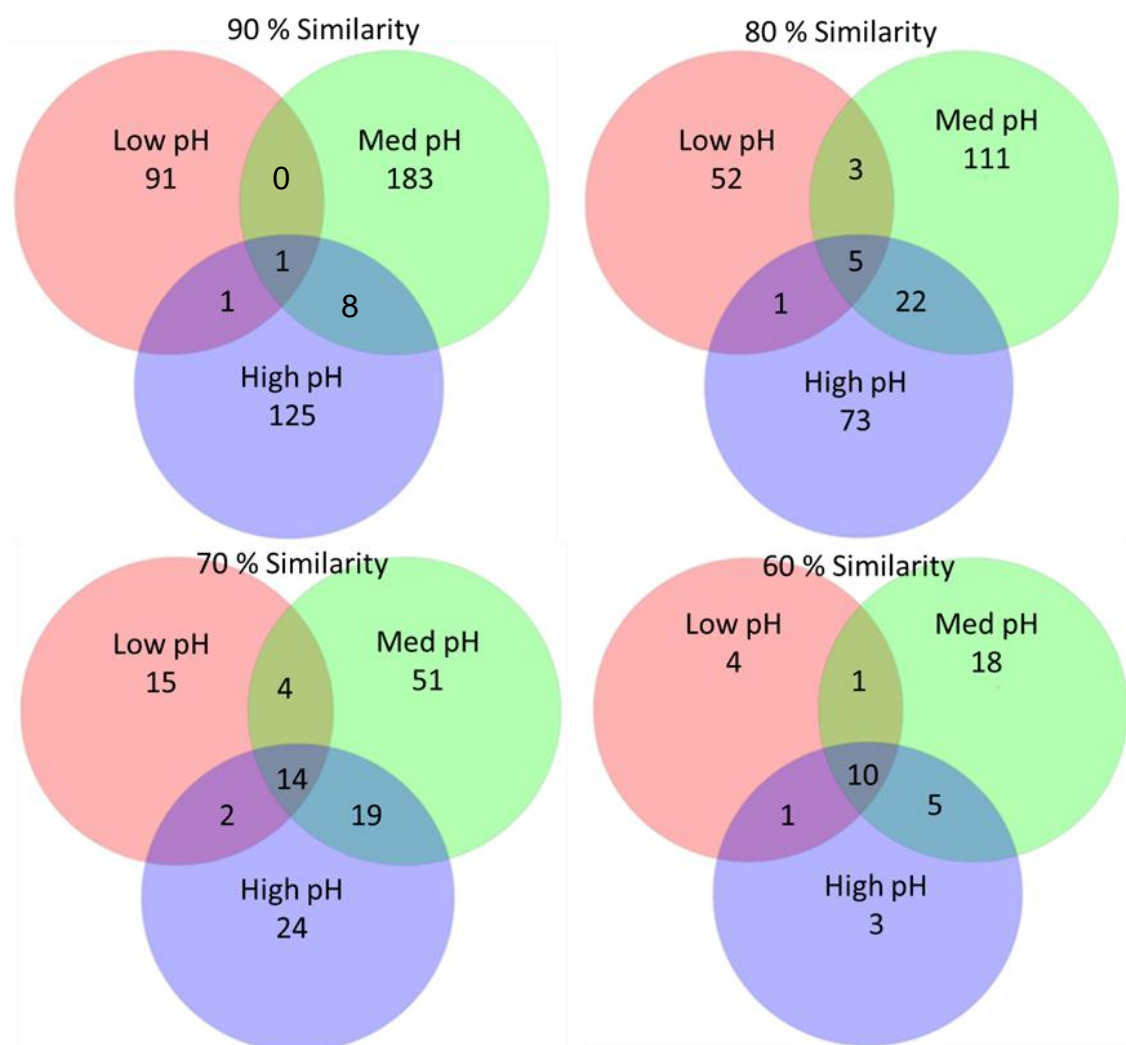


Figure 36: Venn diagrams showing the number of OTUs shared by each of the pH groupings at decreasing levels of OTU similarity. Medium and high pH samples consistently shared more OTUs than either did with the low pH samples

5.3.4 Differences in Community Composition

Differences in community composition were then analysed in each of the samples to examine the effect of environmental variables upon fungal taxa abundance at the class level. A megablast algorithm was used to give identities to each sequence individually by matching sequences against a local quality checked version of the UNITE database at different taxonomic levels. First the changes in the proportional abundances of fungal phyla between and within each pH grouping were examined (Figure 37).

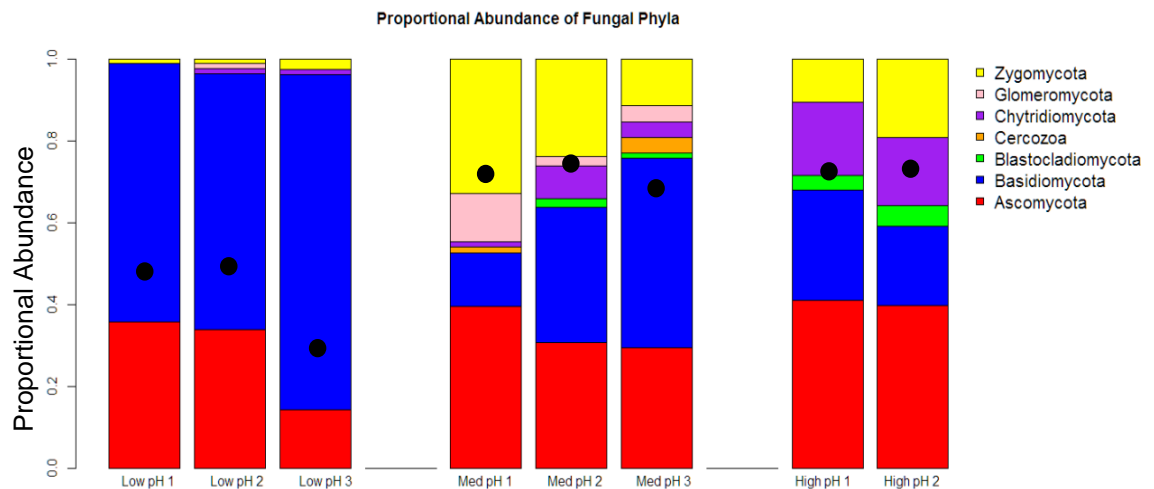


Figure 37: Bar charts showing the community composition of each clone library at the phylum level. Black dots represent the Simpson's diversity for each clone library calculated at the phylum level.

Soils characteristic of low pH environments predominantly contained Basidiomycota, although the proportional abundance of Ascomycota lineages was similar across all samples. Medium and high pH samples had higher proportional abundances of other fungal lineages, and in two medium pH samples protist sequences belonging to the Cercozoa were detected, although at low abundances. Within the high pH samples, Chytridiomycota sequences were detected in higher proportional abundances than in any other group, while Zygomycota sequences were in high proportional abundances within the medium and high pH samples, and were largely absent from the low pH samples. Simpson's diversity index was calculated for each sample based upon the phylum identities and plotted on the graph in the form of black circles as these values are also constrained between 0 and 1. Low pH samples had lower Simpson's diversity scores, whereas high and medium pH samples had similar, high diversities. To analyse these patterns further, bar plots were created to describe the taxonomic composition of each sample at the class level (Figure 38).

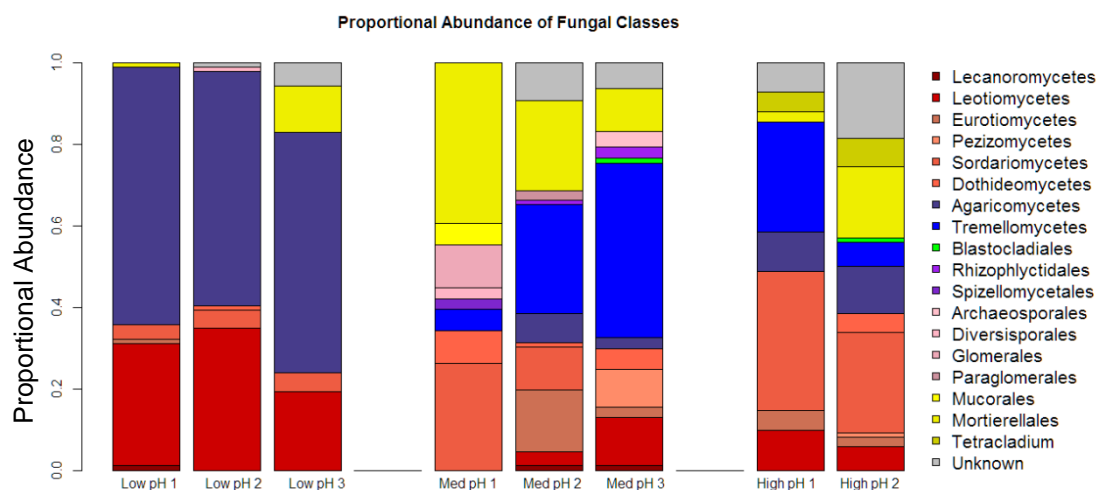


Figure 38: Bar charts showing the community composition of each clone library at the class level. Red bands indicate Ascomycota, blue bands Basidiomycota, purple bands Chytridiomycota, pink bands Glomeromycota and yellow bands Zygomycota.

The Basidiomycota class Agaricomycetes had high proportional abundances in the low pH samples, and few other Basidiomycota lineages were detected. Conversely, within the medium and high pH samples, Agaricomycetes had lower proportional abundances, and the dominant Basidiomycota class was the Tremellomycetes. Ascomycota diversity within the low pH samples was low, with the dominant Ascomycota group being the Leotiomyces. Within the medium pH samples, a range of Ascomycota classes was detected, whereas the high pH samples had a dominance of Sordariomycetes. Zygomycete taxa were largely absent from the low pH samples, while those identified as Mortierellales were seen in higher abundances within the medium pH samples and high pH sample 2. Although Chytridiomycetes were seen in high abundances within the high pH soils in Figure 37, few chytridiomycete classes were detected as class level identifications were not available, hence the rise in sequences with unknown identities in the high pH samples.

In contrast to the t-RFLP based study, clone library analysis suggests that fungal community composition is highly related to the environment. At broad scale taxonomic identifications (phylum – class), samples originating from acidic organic habitats contained fewer distinct fungal phyla and classes, and were dominated by Basidiomycota lineages (predominantly the Agaricomycetes). Conversely, improved habitats with higher pH values were largely dominated by Ascomycota taxa of different classes, and a larger number of fungal phyla and classes were detected at comparatively high proportional abundances.

Accurate higher order taxonomy assignments could not be generated due to the incomplete nature of ITS sequence repositories (Buee *et al.* 2009; Nilsson *et al.* 2006). In order to assess the total sequence variability within and between the different soil environments, a phylogenetic tree was constructed to examine patterns at higher taxonomic resolutions (Figure 39). Samples from the low pH organic environments were highly similar in ITS sequence variability and clustered together in a reduced number of clades compared with the medium and high pH samples. Furthermore, sequences from these samples clustered together, suggesting a greater similarity within samples than between. This further suggests that each low pH organic environment under study contains few, highly related fungal organisms. However due to the dominance of a few fungal taxa, slight differences between these communities result in a large between group variability, as seen in Figure 35. Sequences from medium and high pH soil environments were dispersed across the tree and further highlight the larger numbers of fungal lineages detected within these samples compared with those obtained from the low pH soil environments. Minimal sample clustering was observed for the medium and high pH samples, showing that sequence variability was comparatively large across all samples thus implying higher diversity.

Phylogenetic Placement of Fungal Sequences within a Bootstrapped Neighbour Joining Tree Showing the Phylogenetic Relatedness of Sequences and their Respective pH Groups

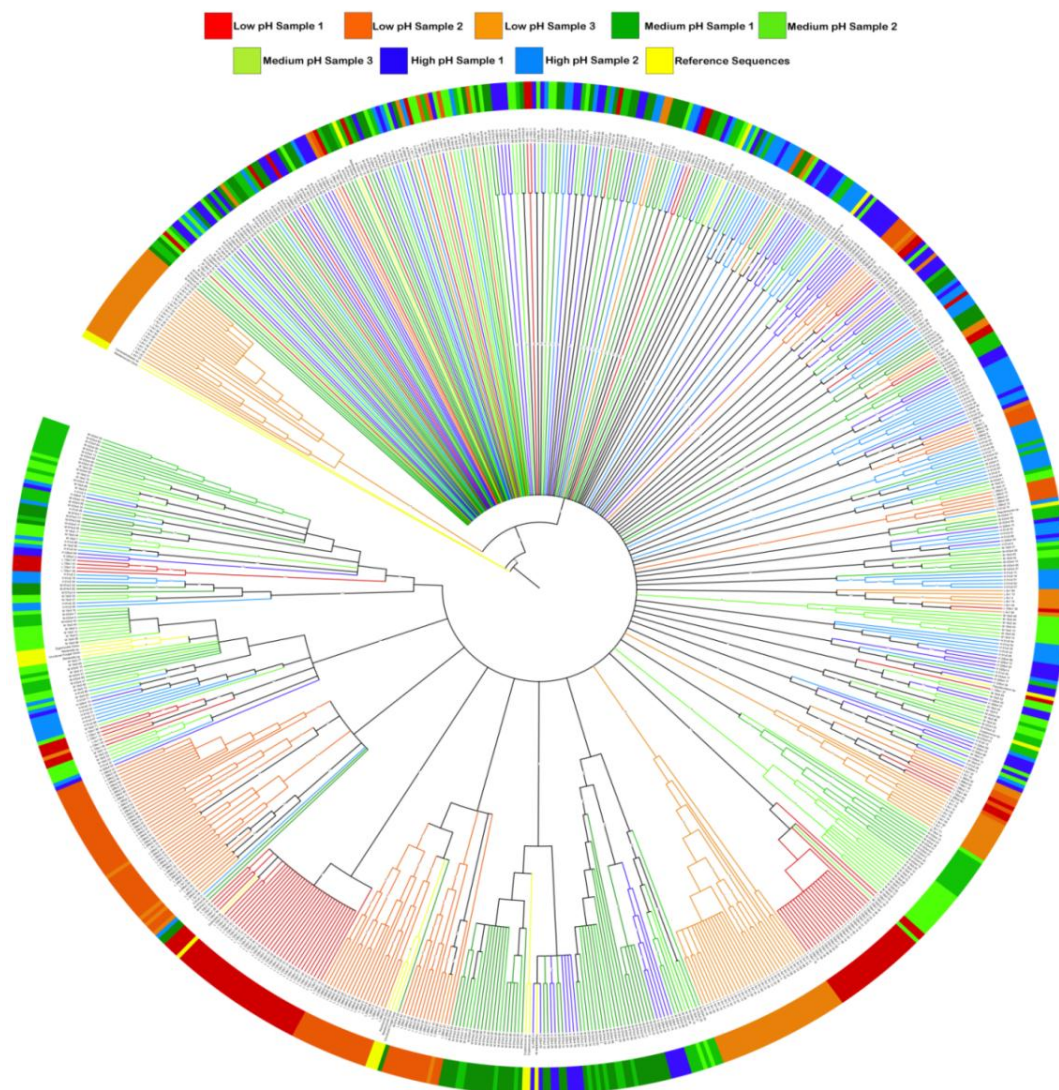


Figure 39: Bootstrapped ($n = 100$) phylogenetic tree showing evolutionary similarity between clones taken from each sample across the pH gradient. Clades with a bootstrap value below 40 were collapsed and ancestral lineages were coloured to denote sample of origin.

5.4 Discussion

The landscape scale assessment of fungal biogeography using this ITS t-RFLP approach indicated that environmental influences upon fungal community structure were minimal. Unlike the reported effects of environmental variables upon bacterial diversity, no definitive links could be made between soil fungal diversity and environmental variables. This suggests that soil fungal organisms are structured differently from other soil microorganisms at the spatial scale and

taxonomic resolution of this study. However, as seen in Chapter 3 with the bacterial based assessment of pH and diversity effects, fungal beta diversity was greater within low pH soils than high pH soils. This highlights the fact that fungal community variability was greater within soils of an acidic organic nature than those closer to neutral or at higher pH values, suggesting that soil pH has a distinct effect upon fungal community variability, again potentially due to the higher levels of environmental variability found within soils of low pH. This suggests that, although distinct patterns in alpha diversity were not seen in relation to environmental variables, fungal communities do respond to variation in soil pH when analysed as meta populations.

Although significant but weak links were seen between fungal community structure and the dominant plant species present at a sample site, the exact nature of this correlation is unclear. The dominant plant species present at a sample site was the strongest predictor of fungal community structure over all other environmental variables. However, other measures of the above ground floristic community, such as the DCA Axis scores of plant community ordinations, had minimal effects. This may be due to the differences in the nature of these variables. The diversity of every sample was high, and evenness was reasonably low (in concurrence with Fierer *et al.* 2007), suggesting that every sample was different from every other sample in terms of the t-RFLP profiles generated. Variables measured in a continuous fashion as a gradient across space (pH, C:N, DCA axis scores, etc.) are unlikely to correlate in a linear fashion with fungal community dissimilarity (across the NMDS axes) due to the high level of between profile variability. However, discrete variables such as the dominant plant species, allow a grouping of profiles based upon minimizing profile-profile dissimilarity. This therefore suggests that, at the spatial scale and taxonomic resolution of the t-RFLP based study, fungal community structure is governed less by large scale ecological gradients and more by localised variation in environmental characteristics. Most variables were found to be significant within the least squares regressions upon the NMDS axes, but the variance explained by each was small.

Studies examining the relationships between soil fungal communities and above ground plant species have found strong links at the local scale, and when targeted towards specific fungal lineages (Mummey *et al.* 2005; van der

Heijden, 1998). Root associated symbiotic relationships between plants and fungal lineages are well documented (Allen *et al.* 1995; Azaizeh *et al.* 1995; Grinstead *et al.* 1982), and it is hypothesised that a large proportion of land plants form mycorrhizal associations with numerous fungal taxa (Malloch *et al.* 1980). However, patterns relating to fungal-plant associations are not uniform across all soil ecosystems. Tropical soil ecosystems, with a high level of plant diversity, have been shown to have 10 – 40 times as many plant hosts to fungal symbionts as temperate soil ecosystems, suggesting that biodiversity patterns of host species and fungal symbionts are not consistent across all soil-scapes (Comas *et al.* 2010), and that other biotic or abiotic factors are important for structuring fungal communities. Temperate ecosystems (such as those examined here) have a high level of plant symbiont diversity, and mycorrhizal fungi were not the only organisms studied here. This may explain why minimal environmental influences upon fungal community structure and diversity were found. It may be the case that the presence and abundance of different fungal lineages are structured in relation to different environmental variables, and landscape scale patterns of biodiversity in relation to environmental variance would become apparent if individual fungal lineages were analysed independently from one another.

Soil nutrient status has been shown to significantly affect the fungal:bacterial biomass ratio in small scale studies. Forest soils with more recalcitrant leaf litters have greater fungal:bacterial biomass ratios than cultivated and grassland soils (Lauber *et al.* 2008). Furthermore, Lauber *et al.* (2008) found that carbon to nitrogen ratio and extractable phosphorus were both correlated with the Ascomycota to Basidiomycota ratios, as soils with higher levels of extractable phosphorus and lower carbon to nitrogen ratios contained more Ascomycota than Basidiomycota. Within the t-RFLP study, there was little evidence to support this finding. A large scale analysis of soil fungi using pyrosequencing showed that a high level of local scale richness exists within fungal populations, and this may be due to a multitude of factors including high levels of micro-scale soil variability, rapid rates of speciation, high immigration rates, low rates of extinction, and small body size (Fierer *et al.* 2007). Furthermore, fungal evenness was far lower than bacterial evenness, a finding supported here, but comparisons between assessments of community structure using different

genes are problematic due to potentially different levels of taxonomic resolution (Fierer *et al.* 2007).

As distinct differences between patterns of fungal and bacterial biodiversity at the landscape scale were seen, the overall community dissimilarity was compared between the two kingdoms using the t-RFLP profiles generated. The level of total community dissimilarity was greater in fungal populations than bacterial populations, further supporting the hypothesis that the number of fungal OTUs exceeds that of bacteria. This hypothesis is, however, confounded by the fact that the level of taxonomic resolution provided by the non-coding ITS region is greater than that of the 16S rRNA gene analysed in Chapters 3 and 4, which highlights the difficulty in making comparisons between microbial kingdoms when using molecular methods to elucidate and compare patterns of biodiversity.

Unlike soil bacteria, pure culture studies of fungal taxa have shown multiple fungal lineages to have a high degree of tolerance to variation in pH (Sletten & Skinner, 1948; Hesse & Ruijter, 2002), and some taxa have optimum growth rates spanning 5 – 9 pH units (Rousk *et al.* 2010). Isolates have also exhibited extreme pH tolerance even when taken from soil environments of approximately neutral pH (Hung, 1983), suggesting a widespread tolerance to extremes of soil pH. A reduction in bacterial diversity within low pH soils has been attributed to an increase in aluminium availability (Pietri & Brookes, 2008), but the toxic effect of aluminium upon fungal communities is not as great as upon bacterial communities (Jones & Muehlchen, 1994). However, although fungal community dissimilarity is largely unaffected solely by soil pH, the ratio of fungal to bacterial biomass has been shown to be determined by soil pH (Rousk *et al.* 2009). Soil fungal biomass is greater than bacterial biomass within low pH soils (potentially due to the reduction in bacterial diversity), and fungal growth rates are also higher in acidic soils. This suggests a competitive advantage for fungi over bacteria, potentially explaining the high levels of fungal diversity seen here across soil pH gradients.

If these findings are compared with what is known about the biogeography of larger terrestrial organisms, similar patterns emerge. A key principle of biogeography is that areas with similar environmental conditions may show

different patterns in biodiversity formed through processes such as limitations to dispersal and competition (Cox & Moore, 2010). As this is clearly the case here, and minimal environmental effects were seen in the t-RFLP based study, it can be hypothesised that fungal populations are structured with greater reference to ecological processes rather than the environmental variability. However, before this can fully proposed, potential limitations to the t-RFLP method used must be assessed.

5.4.1 Limitations to Fungal ITS t-RFLP

A principal limitation to t-RFLP based studies of soil fungal communities is related to the physiology of fungi in general. Ectomycorrhizal fungi form sporocarps that may disseminate spores above and below ground, while hyphal mats and extra radical mycelia form macro structures through the soil matrix. These structures are fragile but relatively ubiquitous and thus the source location of all DNA in field collected samples is difficult to determine (Avis *et al.* 2006). Due to this, DNA from these structures may be incorporated into analysis even if the point of origin is uncertain. Fungal genome structure is another potential problem. The genes coding for ribosomal DNA are numerous and distributed many times within and between chromosomes (Rooney & Ward, 2005). ITS inter species variability is high (Nilsson *et al.* 2008) and intra species variability has been detected in multiple fungal taxa (Avis *et al.* 2005). Furthermore, t-RFLP analysis of single species isolates has produced more t-RFLP peaks than expected (Koide *et al.* 2005) thus artificially inflating richness and diversity estimates. Avis *et al.* (2005) found, using artificial communities, that diversity was consistently over estimated and concluded that ITS based t-RFLP may not give accurate fungal richness estimates.

Assessments of the primers used here suggest that, in conjunction with the restriction enzyme HaeIII, t-RFLP profiles generated by *in silico* digestions of fungal ITS sequences give the most accurate representation of species composition available at the time of publishing (Lord *et al.* 2002). This suggests that the t-RFLP method discriminates between taxa at a relatively high level of taxonomic resolution in comparison to 16S rRNA gene approaches. Upon reflection, this perhaps is not the best approach when analysing landscape scale patterns in fungal biogeography, but is potentially more suited to small

scale studies with a high sampling frequency (henceforth referred to as sampling grain). The t-RFLP profiles generated here show a high level of profile-profile dissimilarity. This may have been due to the high level of taxonomic discrimination available as well as the confounding factors mentioned above. To assess broad scale changes in microbial communities at large spatial scales, a method with lower taxonomic resolution is required as local scale variability in soil edaphic properties is likely to have a greater effect upon species level assessments than those targeted at more broad scale taxonomic groupings (i.e. phylum). Due to the potential problems associated with this t-RFLP method, clone library analysis was undertaken to determine if the absence of large environmental influences upon fungal community structure and diversity was a true finding or the result of methodological limitations.

5.4.2 Clone Library Analysis

In contrast to the large scale t-RFLP assessment of fungal community dissimilarity, clone library analysis showed distinct differences in fungal community composition between independent habitats with different physicochemical properties. However, the high variability of fungal community structure within low pH organic environments was consistent between the two methods. In earlier chapters and current literature, bacterial diversity has been shown to increase across abiotic gradients, most notably soil pH (Fierer & Jackson, 2006; Lauber *et al.* 2009; Chapter 3 - 4), when analysed at broad scale taxonomic resolutions. A similar pattern relating to diversity was seen when organic low pH soils were compared with medium and high pH improved soils using clone library analysis. At lower levels of OTU clustering similarities, Simpson's diversity was lower than that of the medium and high pH soils and total community dissimilarity separated the samples based upon their environmental dissimilarity (i.e. low pH organic, medium pH improved and high pH improved). However, when OTUs were clustered at higher levels of sequence similarity these patterns were not seen. Fungal diversities were always greater than those calculated for bacterial samples, but direct comparisons cannot be made due to potential differences in taxonomic resolution and methodological limitations.

The variability in community dissimilarity within the low pH organic samples was much greater than that of the medium and high pH soils, again showing similarities with findings presented in earlier chapters. In concurrence with Fierer *et al.* (2007) few OTUs were found to be shared between different habitat groupings at stringent levels of OTU clustering, but taxa overlap increased as OTU clustering stringency decreased, and samples with similar environmental characteristics shared more OTUs than those with distinctly different physicochemical properties. As this is in direct contradiction to the t-RFLP based study it is unclear whether fungal populations are indeed structured differently from bacterial populations.

In concurrence with Lauber *et al.* (2008) organic soils with high C:N ratios (here grouped together under low pH soils) were dominated by Basidiomycota taxa. The ratio of Basidiomycota to Ascomycota taxa within these soils was far greater than in the improved soils of the medium and high pH groups. This suggests that, unlike the findings presented in the t-RFLP study, the environment has a strong effect upon fungal community composition. Basidiomycete groups have been linked to the decomposition of low quality (high C:N) lignified substrates (Bardgett & McAlister, 1996), while the anthropogenic improvement of soils (i.e. by fertilisation) has been shown to produce shifts in fungal abundance (Allison *et al.* 2007). This may explain the dominance of Basidiomycota lineages within high carbon, unimproved soils (the acidic organic soils) in comparison with the improved grassland habitats.

At the phylum level, diversity was much lower in the organic acidic soils than in the improved medium and high pH soils. Furthermore the taxa composition across the low pH organic soils was much more even than that of the medium and high pH soils. When community composition was assessed at the class level, Ascomycota and Basidiomycota compositions were different across the 3 groups.

When the ratio of Agaricomycetes to Tremellomycetes was examined across the 3 groups a clear distinction was found. Within the organic low pH soils a dominance of Agaricomycetes was seen in preference to the Tremellomycetes. It is known that many Agaricomycetes produce lignin degrading peroxidases (Morgenstern *et al.* 2008) and thus are highly suited to degrading recalcitrant

plant polymers found at high abundances within organic soils. The Tremellomycetes are a nutritionally and functionally heterogeneous group (Millanes *et al.* 2011) fulfilling a wide range of functional niches including parasitic, fungicolous or parasitic relationships. Due to the non monophyletic nature of this group, analysis with a higher level of taxonomic insight would be needed to assess patterns in biodiversity within this lineage.

The composition of the Ascomycota within the different soil groups was distinctly different. Ascomycota taxa within the low pH environments were predominantly Leotiomycetes, while those in the high pH sample were predominantly Sordariomycetes, a closely related sister taxon (Kurtzman *et al.* 2011). The prevalence of the Sordariomycetes in improved habitats has been shown in other studies (Lauber *et al.* 2008) and is thought to be linked to higher levels of phosphorus availability in improved soils such as grasslands. The Leotiomycetes are the largest group of non lichenised Ascomycota and have wide ranging ecological niches. Little is known about the global distribution of this taxon, and therefore further work may be required to define the environmental limitations of this class (Wang *et al.* 2006).

Chytridiomycota were detected at very low abundances within the low pH soils, as were members of the Zygomycota. Zygomycota were found in higher abundances within neutral pH improved soils, while the Chytridiomycota showed a preference for higher pH improved soils. Few biogeographic studies have examined these fungal lineages, but it is known that members of both phyla may interact with plants through root associations (Read *et al.* 2000) or parasitism (James *et al.* 2006). Few higher order taxonomic identifications could be found for members of this lineage, showing the need for more complete ITS databases if this is to be the gene of choice for fungal community studies.

5.4.3 Conclusions

T-RFLP assessment of landscape scale patterns of fungal biodiversity implies that the relationship between the environment and fungal community structure is distinctly different from that of bacterial populations. However, the above ground plant communities play an important role as determinants of fungal community assembly. Sample-sample variability was high and community evenness low, in concurrence with other studies. The ITS sequence is highly

variable due to the fast rate of mutation accumulation within the non-coding ITS1 and ITS2 regions, and thus potentially discriminates between fungal taxa at a different level of taxonomic resolution from the 16S rRNA gene used in bacterial community profiling studies. This sample variation may obscure community responses to ecological gradients as seen in bacterial studies. However, clone library analysis, where the OTU clustering (and thus taxonomic resolution) can be varied, showed distinct ecological patterns in relation to environmental variables such as soil pH. As seen in the T-RFLP experiment, low pH soils had a greater beta diversity. However, a reduction in alpha diversity and a dominance of specific fungal clades within each pH group were also seen in clone library analysis. This suggests that ecological patterns of fungal biodiversity are somewhat similar to those of bacterial communities, with similar environmental variables being the principal determinants of fungal community assembly.

5.4.4 Further Work

To improve upon the t-RFLP based assessment of landscape scale patterns of fungal biodiversity, a sampling design with a higher sampling frequency or t-RFLP approach with reduced taxonomic resolution could be employed to support or refute findings presented here. An 18S rRNA gene approach may be better suited to this scale of study and show more broad scale patterns rather than those generated at higher taxonomic resolutions. A complete and curated ITS sequence repository is non-existent at present (Nilsson *et al.* 2006), and although the ITS region shows promise for high level taxonomic assignments (Nilsson *et al.* 2008) further work is needed to create a database of sequences without bias towards particular taxa for the wider research community. As clone library and t-RFLP approaches have shown contrasting results within this study, further deep sequencing, encompassing a larger number of samples across distinct environmental gradients and using a different target gene, may provide a better insight into the factors which control and shape soil fungal populations.

Chapter 6: Pyrosequencing Analysis of Soil Fungal Communities

6.1 Introduction

6.1.1 Background

A number of studies have attempted to analyse the environmental conditions that control the distribution and abundance of micro-organisms in the environment (Fierer, 2008; Fierer & Jackson, 2006; Fierer *et al.* 2007; Lauber *et al.* 2009). A preference has been given to determining the ecological drivers of bacterial populations within the soil ecosystem. Fewer studies have targeted fungal communities explicitly with this goal in mind (Anderson & Cairney, 2004; Rousk *et al.* 2010). Fungi are ubiquitous in the soil environment (Zinger *et al.* 2009) and estimates of species numbers range from 1 million to > 15 million (Seifert, 2009), although most mycologists have a “best guess” estimate of 1 million – 1.5 million species (Hawksworth, 2001; Hawksworth, 2004). Soil fungi have many important functions in ecosystem processes; most notably these include nutrient cycling and mediation of decomposition (Zinger *et al.* 2009). There is also evidence to suggest that soil fungi directly affect plant diversity through mutualism or pathogenic interaction (Allen *et al.* 1995; Hattenschwiler *et al.* 2005). This implies that understanding the global distribution of soil fungal populations, and the factors which mediate this distribution, has wide implications for sustainable agricultural practices and soil health in general (Govaerts *et al.* 2008).

Molecular analysis of soil fungal communities can be achieved by several different means. Community fingerprinting methods such as t-RFLP (Lord *et al.* 2002) or estimates of fungal biomass by PLFA (Baath & Anderson, 2003) have been used to examine soil fungal populations. However, these methods are not taxonomically informative as assigning accurate taxonomic identifications to OTUs generated by these methods is troublesome. DNA sequencing based assessments, which allow direct taxonomic identification of OTUs, provide a more detailed insight into changes in community composition across samples of different origin. Clone library analysis is a popular method to generate taxonomically informative insights into soil microbial community composition; however, it is hindered by the small number of sequences commonly obtained per sample. Due to this, only the most dominant organisms within a population

are detected and assessed. Pyrosequencing is currently the most accurate and taxonomically informative procedure available to molecular microbial ecologists, and as such, the unparalleled depth to which this method can delve into complex microbial communities has caused a dramatic increase in its popularity in many aspects of microbial ecology (Kunin & Hugenholtz, 2010).

Unlike for soil bacteria, the relationship between fungal community structure, diversity, and community composition is unclear. Some studies have suggested that soil nutrient status plays an important role in structuring particular fungal taxa, in particular the Sordariomycetes and Agaricales (Lauber *et al.* 2008), whereas others suggest that plant species (and potentially more directly, the resultant root exudate compounds) are the main determinants, at least within root associated soils (Cullings & Hanely, 2010; Mitchel *et al.* 2011; Kaiser *et al.* 2010). Within studies of large scale bacterial biodiversity patterns, soil pH was shown to be a major determinant of diversity and community structure at the landscape scale. However, when fungal communities were analysed across a small scale anthropogenically manipulated pH gradient using an 18S rRNA targeted pyrosequencing approach, pH was shown to have minimal influence upon fungal community structure and diversity (Rousk *et al.* 2010). Soil fungal populations are thought to be highly diverse, with fungal OTUs equalling or exceeding those of bacteria, while evenness scores are substantially lower, implying high variability in total fungal community structure between soils (Fierer *et al.* 2007). Pyrosequencing analysis of six different forest soils using an ITS targeted pyrosequencing approach supports this conclusion (Buée *et al.* 2009), showing that although no consensus has been reached regarding the optimum gene for use with soil fungal pyrosequencing assessments, similar patterns in diversity and evenness are seen when different genes are used for fungal community analysis. However, ITS based pyrosequencing approaches, although theoretically able to discriminate between closely related fungal taxa (Nilsson *et al.* 2008), are hampered by the absence of a large, curated repository of identified gene fragments (Nilsson *et al.* 2006). Due to this, ITS pyrosequencing studies often return large numbers of unidentified taxa, as highlighted in Buée *et al.*, (2009). General trends suggest that Basidiomycota are dominant within low pH organic soils (Lauber *et al.* 2008; Buée *et al.* 2009;

Chapter 5), but in depth assessments of soil fungal community structure across large geographic scales are not prevalent within current literature.

The previous chapter assessed landscape scale patterns in biodiversity using an ITS targeted t-RFLP approach in combination with a smaller scale clone library sequencing approach. Minimal environmental influences upon fungal community dissimilarity and diversity were found using the t-RFLP approach. In contrast, the clone library analysis showed contradictory results suggesting that samples taken from highly divergent soil systems exhibit distinct patterns in fungal community composition. To build upon these findings, and to provide evidence to support or refute the conclusions presented in Chapter 5, a pyrosequencing approach was undertaken to analyse 15 soil samples from 3 contrasting soil environments largely dictated by soil pH. An 18S rRNA approach was proposed as large 18S rRNA curated sequence repositories are available (Pruesse *et al.* 2007), which provide accurate taxonomic assignments to approximately the family level (Anderson *et al.* 2003). Furthermore, the 18S rRNA gene was chosen to avoid problems associated with high intra-species variability commonly associated with ITS based community assessment approaches (Bergerow *et al.* 2010). Although 18S rRNA sequencing is unable to provide detailed high level taxonomic assignments, at present broad scale patterns in fungal biodiversity are elusive and as such phylum/class assessments are first needed to provide a general insight into patterns of soil fungal biodiversity.

6.1.2 Aims

To support or refute findings presented in Chapter 5 an 18S rRNA targeted pyrosequencing approach, analysing 15 geographically isolated soil samples taken from 3 areas with distinct soil physicochemical properties, was implemented. Multivariate analysis was employed to determine the relative effects of environmental variables upon the total community dissimilarity to highlight linkages between fungal community structure and the environment. Community composition was compared across all samples to highlight phylum and class level differences between soils of different physicochemical composition.

6.2 Materials and Methods

6.2.1 Sample Selection and Nucleic Acid Extraction

Fifteen geographically isolated soils spanning a natural pH gradient (low pH, 4.3 +/- 0.23 (n = 5); medium pH, 6.15 +/- 0.08 (n = 5); and high pH, 8.28 +/- 0.16 (n = 5)) were used in this study (detailed in Chapter 3). For sample locations see Appendix B. All samples within each pH group were chosen as being representative of soils typically found within the UK. Low pH soils were predominantly organic in nature, had minimal anthropogenic manipulation and supported floristic communities typical of soils with these properties. High pH samples were predominantly from improved habitats such as agricultural land or improved grasslands and had distinctly different floristic communities to the low pH samples. Medium pH samples were chosen from a mixture of soil and habitat types in which soil pH had intermediate values compared with the low and high pH samples.

6.2.2 Massively Parallel Fungal Tag Encoded FLX Amplicon Pyrosequencing (F-TEFAP)

All pyrosequencing steps were carried out using optimised protocols by Research and Testing Laboratory (Lubbock, Texas) in the same manner as detailed in section 4.2.3. Fungal specific primers 515 F (5'- GTG CCA GCM GCC GCG GTA A -3') and 1100R (5'- TCG GCA TAG TTT ATG GTT AAG -3') were used to amplify an approximately 585 bp length of the 18S rRNA gene.

6.2.3 Sequence Processing

Sequence processing was carried out in the same fashion as described in section 4.2.4.

6.2.4 Data Analysis

Uclust (Edgar, 2010) was used to cluster sequences into OTUs at 97 % similarity. Seed sequences for each OTU were used to represent each OTU, and taxonomies were assigned by retrieving the top BLAST hit for each sequence with a maximum e-value of 0.001. The reference database by which BLAST assigned taxonomy was the generic taxonomy mapping database supplied by SILVA (Pruesse *et al.* 2007) and recommended by the

documentation in the Qiime software package (Caporaso *et al.* 2010). OTU tables were split upon the basis of assigned taxonomy at the kingdom level so that only fungal sequences were used in subsequent analyses. Alignments were generated using ClustalW (Thompson *et al.* 1994) and filtered for common vertical gaps. Chimeraslayer (Haas *et al.* 2011) was used to screen for chimeric sequences against the same database used to assign taxonomy.

Principal co-ordinate analysis (PCoA) was used to visualise jackknifed ($n = 10$) un-weighted and weighted Unifrac (Lozupone *et al.* 2011) scores for each sample. Hierarchical clustering using the un-weighted pair group method with arithmetic mean (UPGMA) was also implemented to visualise the clustering of each sample across the pH classes. Jackknife support ($n = 10$) was utilised by comparing UPGMA trees constructed from a subset of the sequences (75 % of the smallest sample) with the tree created from all sequences, multiple times ($n = 10$). Nodes of the resultant tree were coloured to show Jackknife support (red = 75 – 100 %, yellow = 50 – 75 % and green 25 – 50 %).

An NMDS ordination was constructed from the OTU tables generated in Qiime to which environmental variables were fitted in the form of a least squares regression to determine which environmental variables were responsible for the differences in community composition between samples.

Rarefied alpha diversity metrics (principally Chao1) were computed using the Qiime pipeline (Caporaso *et al.* 2010), and relative abundances were calculated for various levels of taxonomic classifications (phylum – family where possible). All graphics and subsequent data analysis steps were carried out using R (Ihaka & Gentleman, 1996) with the packages *vegan* (Oksanen, 2011) and *BioDiversityR* (Kindt *et al.* 2005).

FastTree (Price *et al.* 2010) was used to construct phylogenetic trees from the sequence alignments after the 10 % most variable positions of the sequences were removed, as were positions that exhibited more than 80 % gaps. The interactive tree of life (ITOL) (Letunic & Bork, 2011) was used to visualise phylogenetic trees. Taxa assignments and bar charts representing the mean proportional abundance of each OTU across the pH groups were added to terminal nodes.

Indicator taxa analysis was undertaken using the function `indval` in the R package `labDSV` (Roberts, 2010) and was performed in order to provide statistical links between individual fungal taxa and the pH group of origin.

6.3 Results

6.3.1 Sequence processing

Table 10 shows the number of sequences returned after sequence processing, and the number of OTUs generated from the processed sequences.

Sequences post filtering stage	Low pH samples 1 - 5				
Pre filtering	7641	12904	6277	13311	10055
Quality	4015	5864	2977	5651	4176
Length	3318	4414	2377	4703	3388
Homopolymer	3301	4397	2300	4652	3298
Primer Mismatch	3105	2262	1368	2767	1653
Chimeras	3100	2259	1367	2757	1652
OTUs at 97 % Similarity	94	44	74	84	88
	Medium pH Samples 1 - 5				
Pre filtering	12290	11947	9895	10485	13392
Quality	5372	5213	4412	5533	6244
Length	4304	4184	3453	4519	5207
Homopolymer	4293	4179	3448	4502	4204
Primer Mismatch	2379	1890	1417	2851	3632
Chimeras	2374	1881	1409	2846	3630
OTUs at 97 % Similarity	245	289	212	64	333
	High pH Samples 1 – 5				
Pre filtering	12324	6783	13950	8565	9767
Quality	9390	2830	6455	3869	4120
Length	5229	1842	4803	2755	3417
Homopolymer	5211	1823	4769	2743	3414
Primer Mismatch	4789	1572	2281	1351	1487
Chimeras	4783	1560	2279	1347	1486
OTUs at 97 % Similarity	214	66	220	152	193

Table 10: Number of fungal sequences pre and post filtering for each sample, and the calculated number of OTUs in each sample at 97 % sequence similarity.

6.3.2 Phylogenetic Estimates of Soil Fungal Community Dissimilarity

PCoA of Unifrac distance matrices was performed to visualize fungal community structure differences between samples from soils of different pH (Figure 40). When 18S rRNA gene sequences were used to calculate a Unifrac based dissimilarity matrix, a distinct separation of the samples based upon pH group was observed. Inter quartile ranges of samples from separate pH groups did not overlap, suggesting a lower level of within pH group variability than between pH group variability.

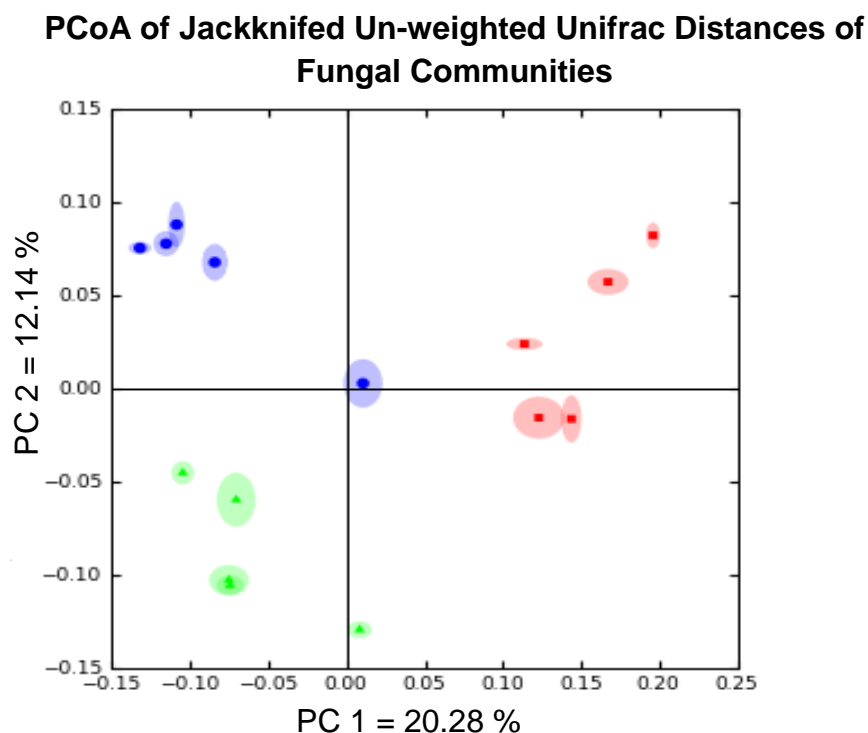


Figure 40: Principal co-ordinate analysis of Jackknifed ($n = 10$) un-weighted Unifrac distances of soil fungal communities across a pH gradient. Red = low pH, green = medium pH, and blue = high pH. Ellipsoids represent the inter quartile confidence ranges.

In order to give a visual representation of the within and between pH group variability, hierarchical clustering (un-weighted pair group method with arithmetic mean) was performed (Figure 41)

UPGMA Hierarchical Clustering of Low, Medium and High pH Samples Based upon Soil Fungal Community Dissimilarity

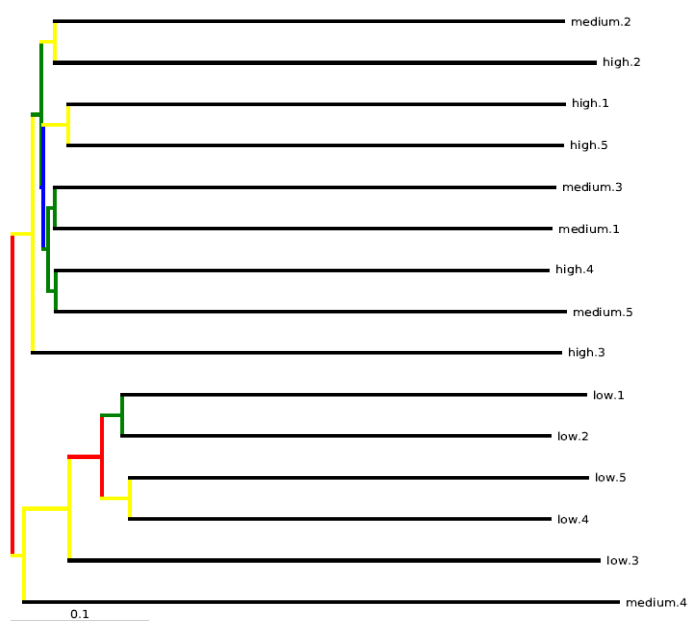


Figure 41: UPGMA hierarchical cluster analysis of low, medium and high pH soil types with jackknifed ($n = 10$) confidence assigned by colour to each branch. Branches which split at the same position between 75 % and 100 % of the sub trees generated for Jackknife support are coloured red, 50 % - 75 % are coloured yellow, 25 % - 50 % are coloured green and 0 % - 25 % are coloured blue

Low pH samples clustered together with strong Jackknife support. However one medium pH sample (medium.4) clustered loosely with the low pH samples, suggesting a more similar community structure (in phylogenetic terms) than that of the other medium and high pH samples. Medium and high pH samples, on the whole, clustered with one another but away from the low pH samples. This suggests that the fungal communities found in acidic soils are different to those found in medium and high pH soils.

To determine which environmental variables had the greatest effect upon fungal community structure, an NMDS ordination was constructed from the OTU table generated at 97 % sequence similarity. To this, environmental variables were fitted via a least squares regression and general additive models (Figure 42). The DCA Axis 1 scores representing the above ground plant communities were responsible for the large changes in fungal community structure along both NMDS axes, along with soil pH ($R^2 = 0.80$ and 0.79 respectively, all $p < 0.001$). Soil moisture and loss on ignition also explained a large proportion of the variance within the plot ($R^2 = 0.72$ and 0.65 respectively, $p < 0.001$). However, all other environmental variables were less significant as p values exceeded

0.001. As a high degree of confidence was desired, further analysis of lower confidence variables was abandoned. This shows that fungal community structure is largely governed by environmental variation, with particular emphasis on the above ground plant communities, soil pH, soil moisture, and loss on ignition.

NMDS Ordination of Fungal OTUs at 97 % Similarity

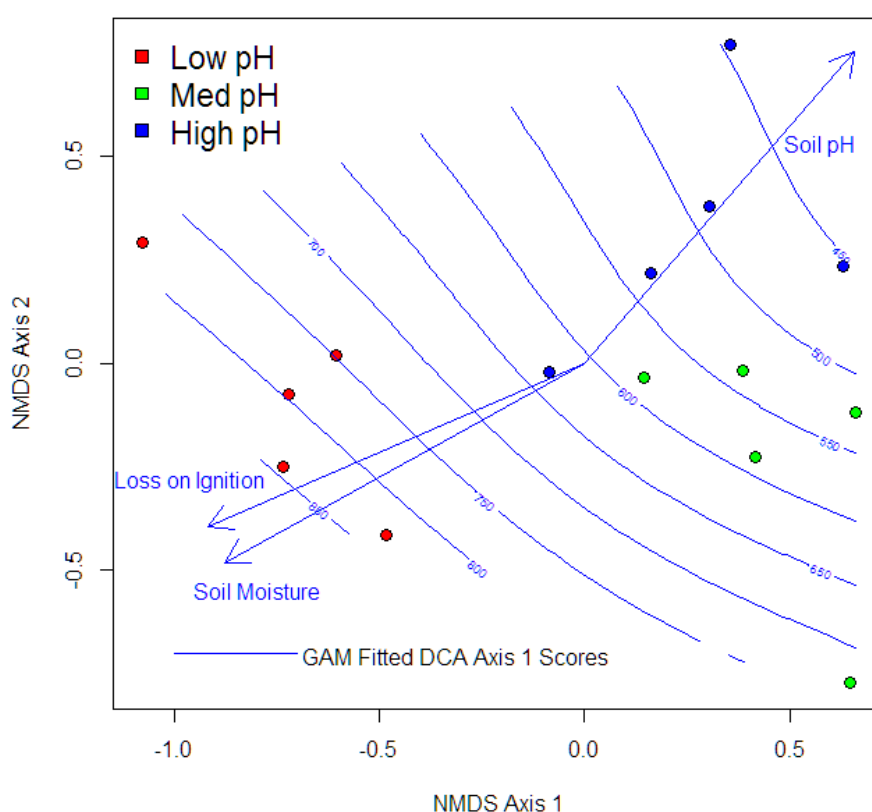


Figure 42: NMDS ordination showing the community dissimilarity of soil fungal populations is largely determined by the above ground plant communities, soil pH, loss on ignition, and soil moisture as defined by a least squares regression fit to the NMDS ordination.

6.3.3 Alpha Diversity

As distinct groupings of samples were seen upon the basis of soil pH and total above ground plant communities, diversity (Chao1 evenness) was calculated for subsamples with equal sequence numbers and averaged within the different groups. Fungal diversity was less within the low pH soil samples as defined by rarefied Chao1 scores. Medium and high pH samples exhibited similar Chao1 scores for each subsample of the sequence populations selected for rarefaction. This is in line with the phylogenetically based assessments of community dissimilarity reported above. The formation of an asymptote was observed within the low pH communities, suggesting a larger proportion of the

diversity was captured compared with medium and high pH samples (Figure 43).

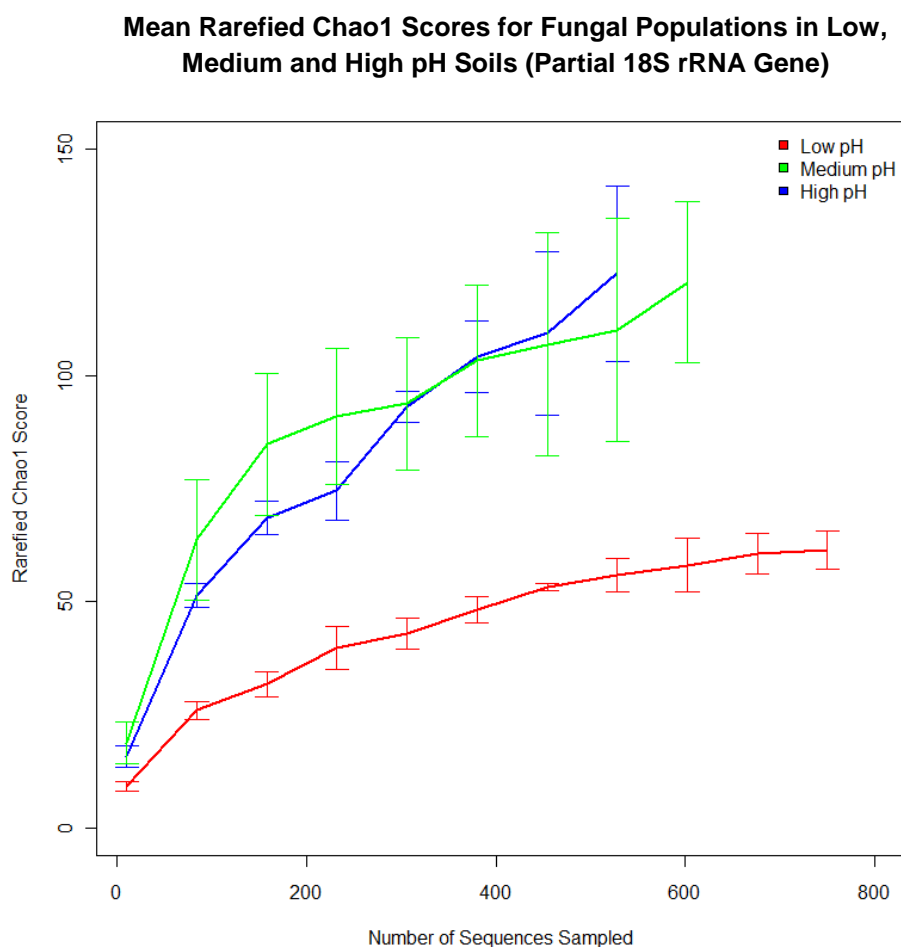


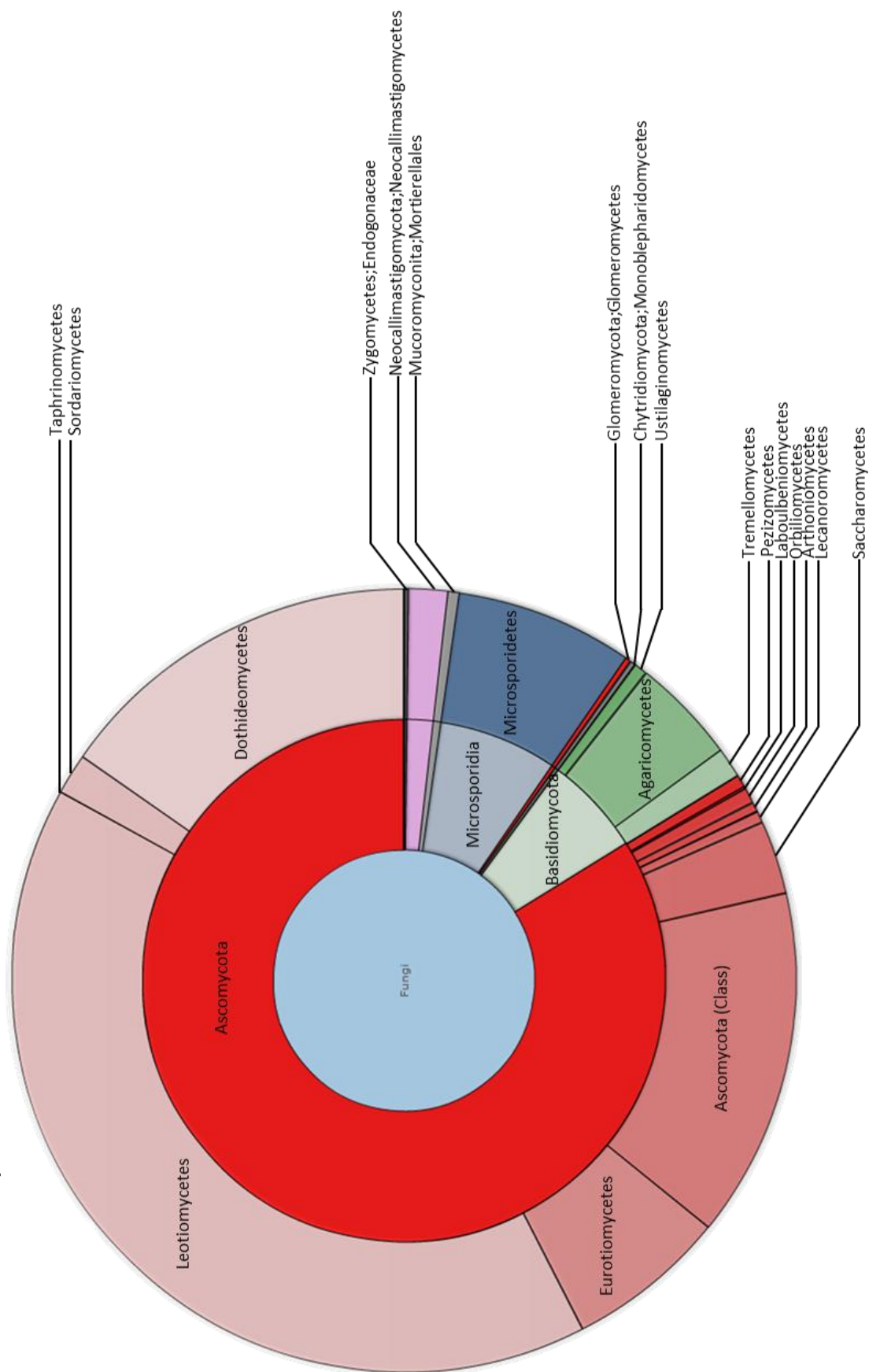
Figure 43: Mean rarefied Chao1 scores for fungal communities residing within low, medium and high pH soils. Error bars denote the standard error of the mean for each rarefied subsample.

6.3.4 Taxonomic Composition

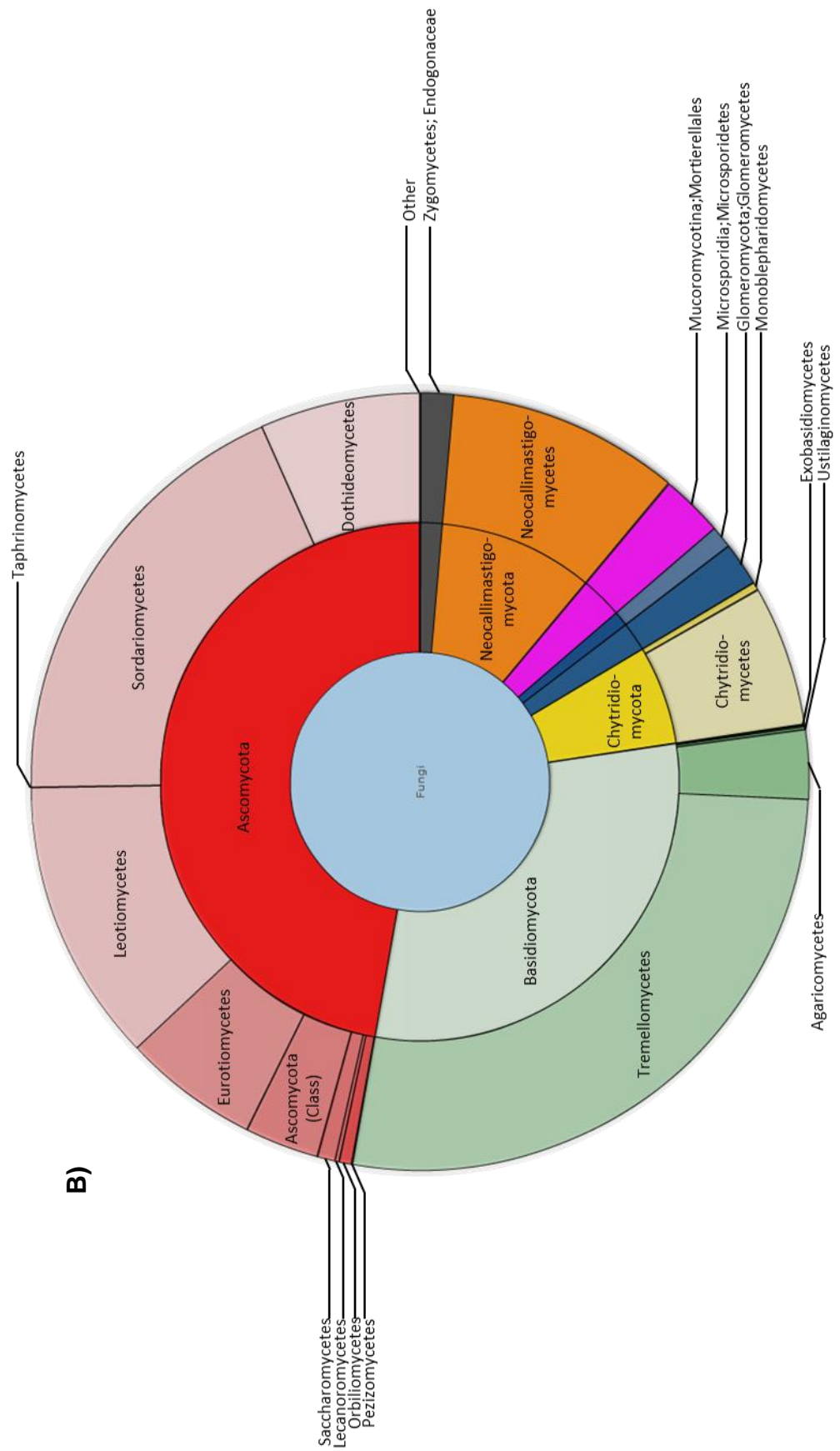
To determine differences in fungal taxonomic composition within the low, medium and high pH soils, the mean proportional abundance of each OTU at the phylum and class level was plotted (Figure 44). Appendix R shows the results per sample to indicate the sample variability of each class between samples.

Mean Proportional abundance of Fungal Phyla and Classes in Low pH Soils

A)



Mean Proportional abundance of Fungal Phyla and Classes in Medium pH Soils



Mean Proportional abundance of Fungal Phyla and Classes in High pH Soils

C)

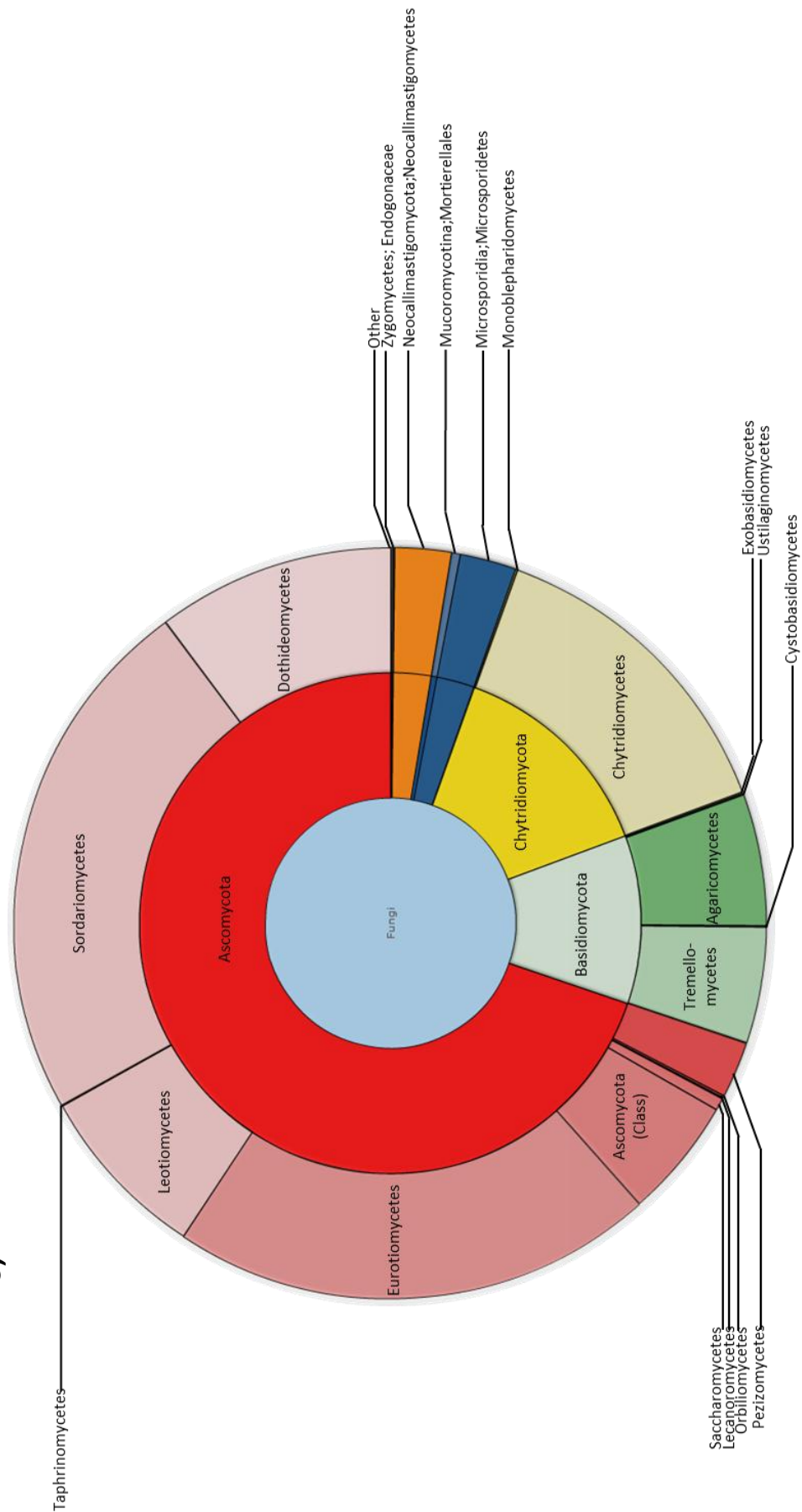


Figure 44: Mean Proportional abundance of fungal phyla (Inner ring) and classes (outer ring) in low pH (A), medium pH (B) and high pH (C) soil samples.

6.3.5 Indicator Taxa Analysis

Indicator taxa analysis was undertaken to provide a statistical representation of fungal classes found predominantly within soils of a single pH grouping and largely absent from soils of other pH classes. Indicator values are used as a metric to represent a combination of taxa relative abundance and the relative frequency of occurrence of each taxon within the various groups of site (Dufrene & Legendre 1996), in this case defined by soil pH. Significance values (probability) are computed using a randomisation procedure as detailed in Dufrene & Legendre (1996). Within the low pH samples the Ascomycota lineages Laboulbeniomyces, Arthoniomyces, and Leotiomyces were defined as indicator taxa. The medium pH soils were characterised by a more diverse set of fungal lineages including those belonging to the Glomeromycota, Zygomycetes, Neocallimastigomycota, and Tremellomycetes. High pH samples were statistically linked with higher abundances of the Chytridiomycota lineage Chytridiomyces. All indicator values were high, showing that these lineages were predominantly found in all samples of single pH groups and predominantly absent from samples of other pH groups (Table 11).

Indicator Taxa	pH Group	Indicator Value (%)	Probability
Laboulbeniomyces	Low	0.80	0.01
Arthoniomyces	Low	0.75	0.01
Leotiomyces	Low	0.67	0.016
Glomeromycetes	Medium	0.89	0.008
Endogonaceae	Medium	0.83	0.01
Tremellomycetes	Medium	0.81	0.03
Neocallimastigomycetes	Medium	0.70	0.02
Chytridiomyces	High	0.69	0.03

Table 11: Indicator taxa analysis displaying fungal taxa statistically linked to a particular soil pH group.

Phylogenetic trees were constructed for OTUs with assigned taxonomies belonging to the Ascomycota, Basidiomycota and Chytridiomycota as these were the most dominant fungal lineages present. Figures 45 to 47 show the phylogenetic placement of members of these taxa and the mean proportional abundance of each OTU within the low, medium and high pH samples. High resolution zoom-able digital copies of each tree are also provided. At 97 % sequence similarity, few OTUs were shared between pH groups. A large proportion of Ascomycota OTUs were unique to the low pH soils and clustered together within the tree, suggesting a close phylogenetic relationship. Those

that had higher level taxonomic assignments belonged to the Leotiomycetes, providing further evidence that this lineage is predominant within low pH organic soils. Leotiomycetes were not solely restricted to the low pH organic soils but few OTUs defined as belonging to these taxa were shared across more than 1 pH grouping. The majority of Basidiomycota OTUs observed within a low pH sample belonged to the Agaricomycetes. However, members of this lineage were also found in medium and high pH soils. Tremellomycetes OTUs were often found within the medium and high pH samples but rarely within the low pH samples and again a minimal number of shared OTUs between pH groups was observed. The Chytridiomycetes were predominantly found in the high and medium pH samples, and only three OTUs were observed in the low pH samples. No distinct pattern was seen in taxa unique to a pH group, but in conjunction with abundance plots and indicator taxa analysis, it can be hypothesised that the Chytridiomycetes may have a defined niche within high pH soil systems. In conclusion, low pH soils are characterised by a high proportional abundance of Ascomycota, particularly the Leotiomycetes, while medium and high pH samples are characterised by Basidiomycota and Chytridiomycota lineages. No taxonomic assignments were wholly unique to a single pH grouping at the phylum or class level, therefore further work may be needed to analyse patterns in fungal biodiversity closer to the species level.

When OTUs were clustered at 97 % similarity, a high degree of OTU endemism was observed within each pH grouping. Using all fungal sequences, a radial tree representing OTUs clustered at 90 % similarity was constructed and heat maps representing the presence or absence of each OTU across all samples were added to the terminal nodes (Figure 49). Again, few fungal OTUs were shared across all samples, and many were unique to one or a few replicates of each pH class. Most notably, Chytridiomycota taxa were present in most medium and high pH samples but largely absent from low pH samples. Certain Ascomycota taxa, with low level taxonomic identifications, were unique to most replicates of the low pH samples. Basidiomycota OTUs were predominantly found within many of the medium pH samples and a limited number of the high pH samples. This indicates that all fungal OTUs, at this level of de-replication or sequencing depth, are not found ubiquitously across soil biomes and that environmental influences highlighted earlier (predominantly above ground plant

community and soil pH) affect patterns of soil fungal biodiversity. Further research may be required to analyse shared and unique fungal OTUs by encompassing a larger number of soil habitats, and at greater levels of sequencing depth.

Phylogenetic Tree of Mean Proportional Abundance of Ascomycota OTUs in Low, Medium and High pH Soil Samples

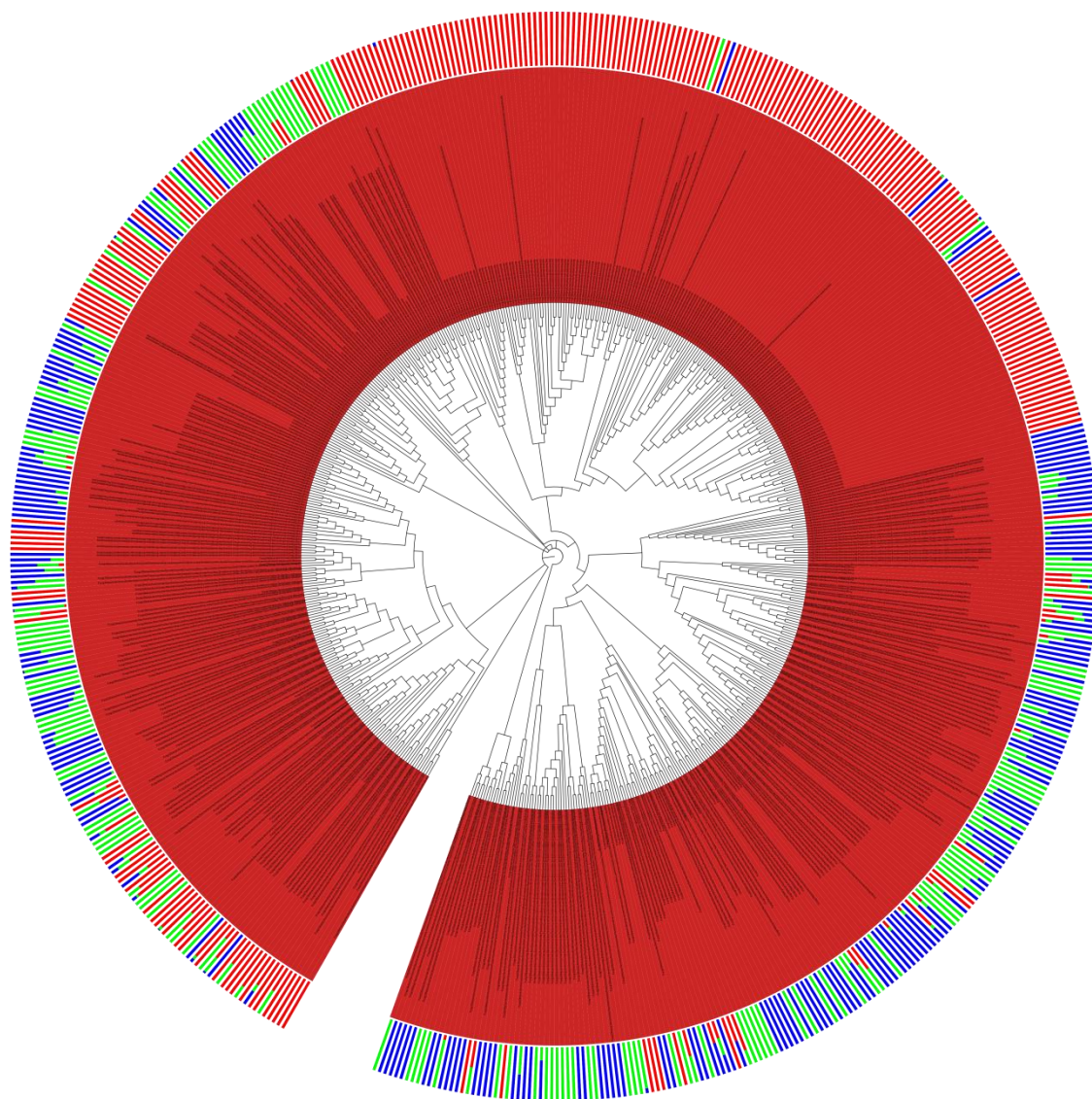


Figure 45: Phylogenetic placements of Ascomycota OTUs and the mean proportional abundance of each OTU in low (red), medium (green) and high (blue) pH soil samples denoted by bar charts at each terminal node. High resolution digital images are available on the accompanying CD. Here a number of Ascomycota OTUs were unique to the low pH environments. Many of these were unclassifiable below the phylum level but clustered with the Pezizomycotina lineages Leotiomyces and their closely related sister taxon the Dothideomycetes. Other Ascomycota lineages were seen in high abundances in the medium and high pH samples, many of which also belonged to the Leotiomyces and Dothideomycetes.

Phylogenetic Tree of Mean Proportional Abundance of Basidiomycota OTUs in Low, Medium and High pH Soil Samples

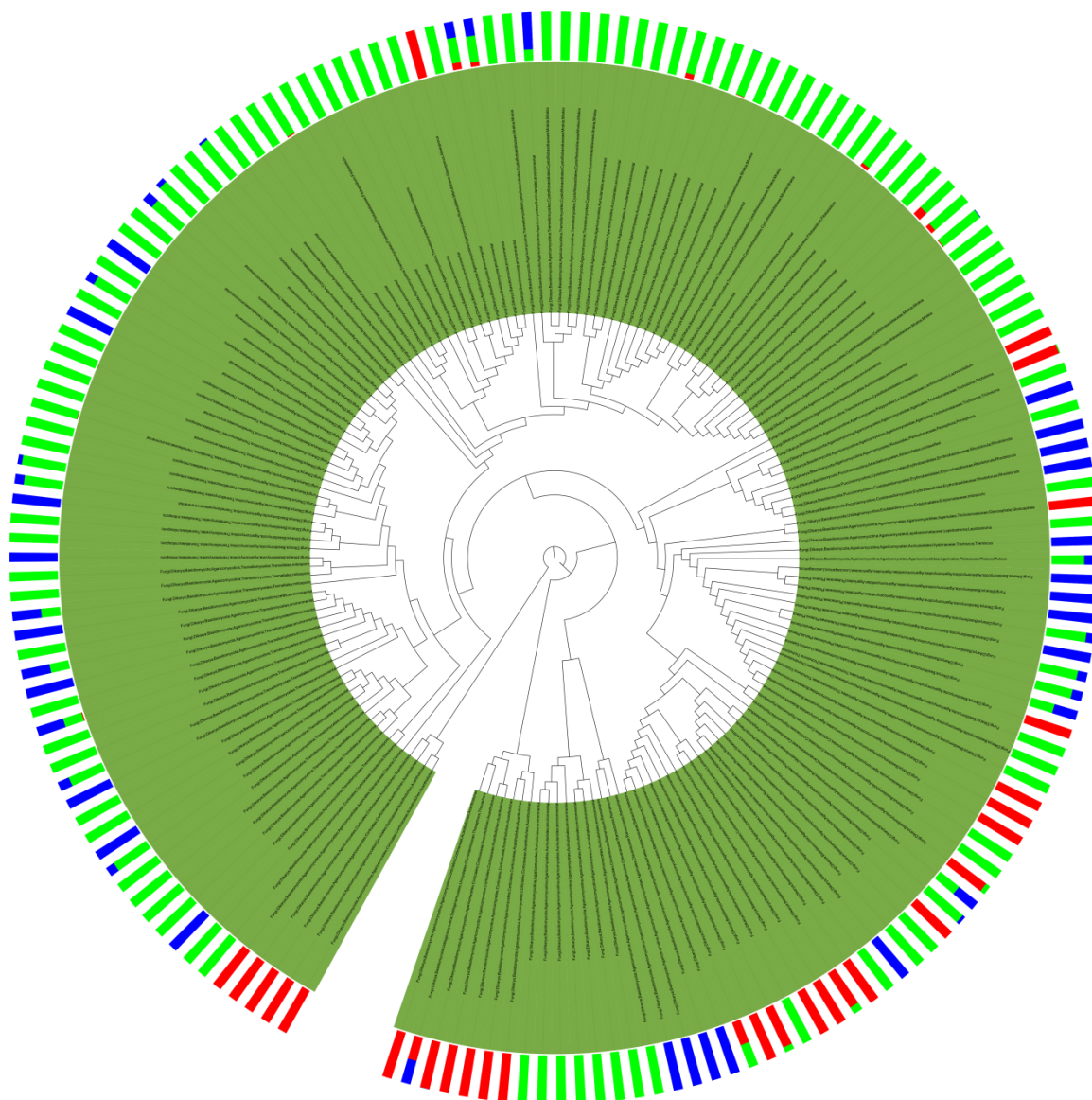


Figure 46: Phylogenetic placement of Basidiomycota OTUs as defined by partial 18S rRNA gene sequences and the mean proportional abundance of each OTU in low (red), medium (green) and high (blue) pH soil samples denoted by bar charts at each terminal node. Many OTUs defined as Basidiomycota were prevalent within medium and high pH samples and comparatively few were detected within low pH samples. In general a large proportion of the OTUs seen in the medium pH samples were of the order Tremellomycetes and Agaricomycotina. Those found within the low pH samples were almost entirely members of the Corticiales order.

Phylogenetic Tree of Mean Proportional Abundance of Chytridiomycota OTUs in Low, Medium and High pH Soil Samples

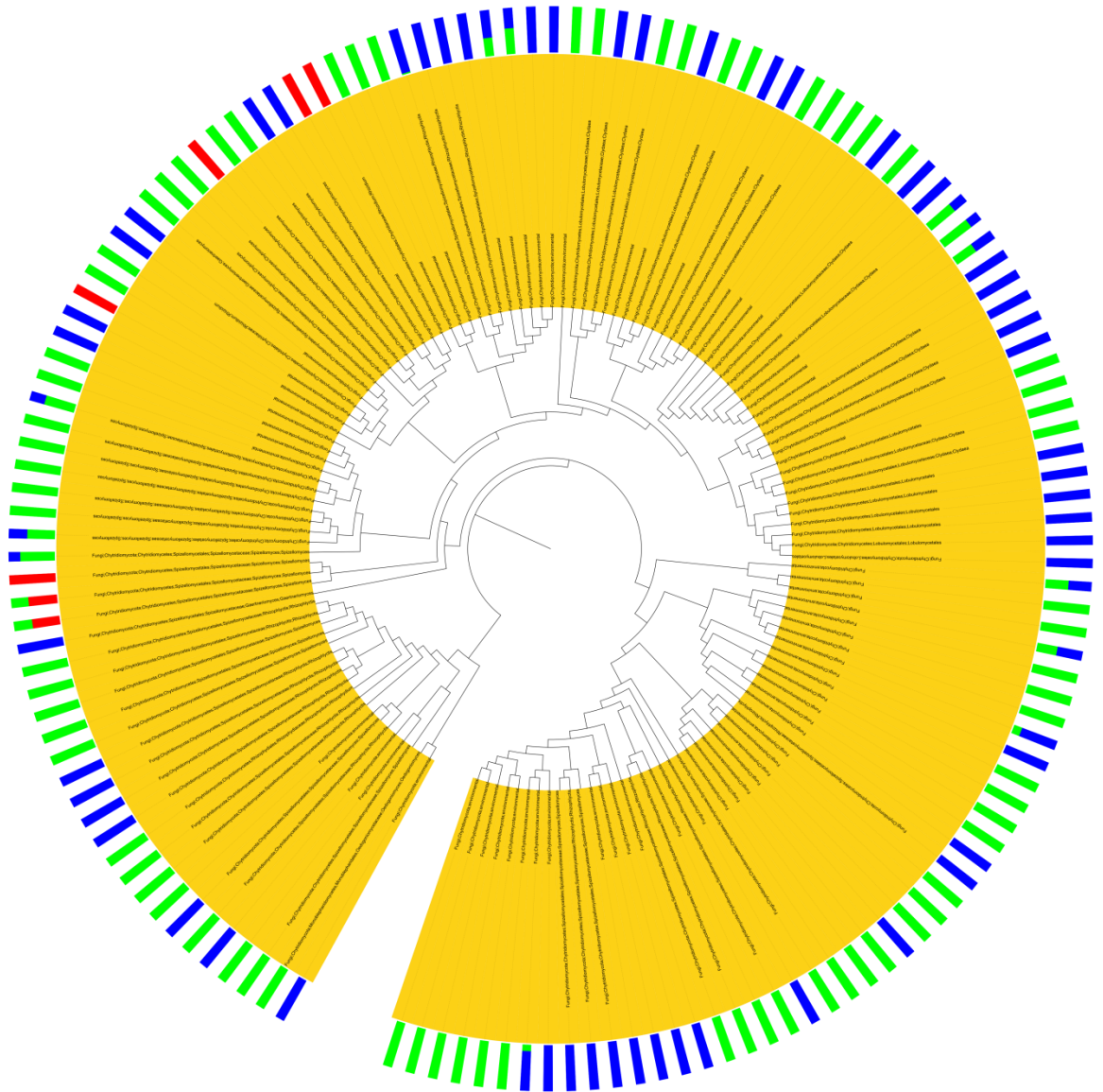


Figure 47: Phylogenetic positions of the Chytridiomycota lineages using partial 18S rRNA gene pyrosequencing. In general Chytridiomycota lineages showed a preference for medium and high pH soils within this study, and most OTUs detected were unique to a particular pH grouping. Bar charts denote the proportional abundance of each OTU within the low (red), medium (green), and high (blue) pH samples.

Phylogenetic Tree of Shared Fungal OTUs Between and Within pH Groupings

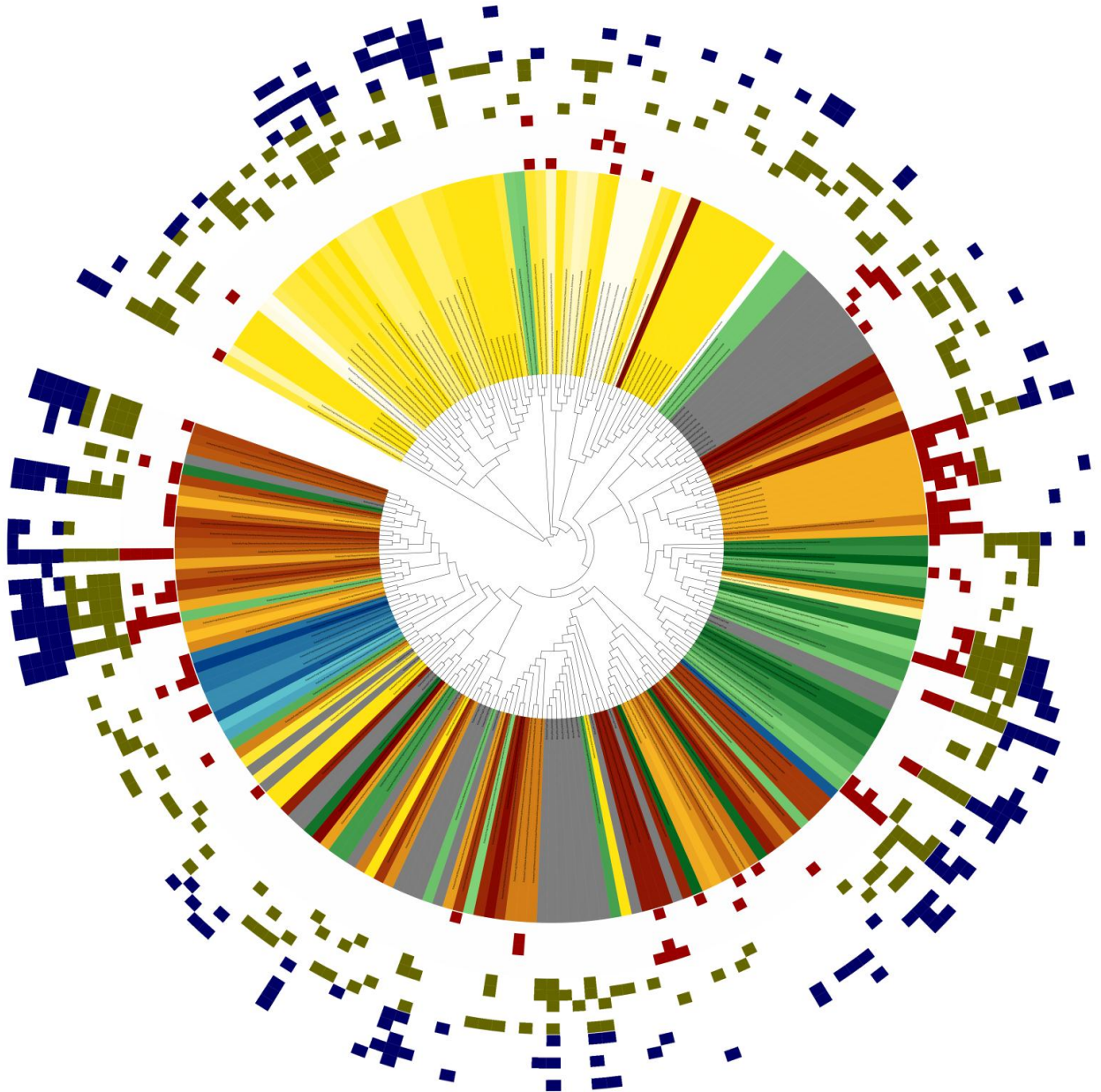


Figure 48: Phylogenetic tree showing the relative positions of fungal OTUs clustered at 90 % similarity derived from pyrosequencing of 15 soils spanning a natural pH gradient. Heat maps at the terminal nodes represent a presence or absence of each OTU within the 5 low pH samples (red), 5 medium pH samples (green) and 5 high pH samples (blue).

6.4 Discussion

In concurrence with the clone library assessment of fungal biodiversity presented in Chapter 5, fungal community structure, composition, and diversity had strong links with the environment. The above ground plant species and soil pH were highlighted as potentially important predictors of fungal community structure within this study, as they were in the bacterial pyrosequencing assessment, although R^2 values were lower compared with the results in Chapter 5. This suggests other factors not assessed here also play significant roles in determining fungal community structure. Loss on ignition (indicative of organic matter content) and soil moisture were also significant within the model, although they were less important than the dominant plant species or soil pH as determinants of fungal community structure. This shows that some of the factors known to affect the community structure of other soil microbial populations (Fierer *et al.* 2007; Lauber *et al.* 2009; Chapter 3 - 4) have similar effects upon fungal community structure.

6.4.1 Changes in Fungal Communities Across Environmental Gradients

Using phylogenetically based dissimilarity measures, the variability of total fungal community structure within and between defined groups of samples can be assessed (Caporaso *et al.* 2011). While all three pH groups were seen to be confined to distinct areas of the PCoA plot of Unifrac dissimilarity matrices, UPGMA trees suggest that fungal community composition of low pH samples is substantially different from the medium and high pH samples, while the medium and high pH samples are reasonably similar in terms of fungal community composition. This may be predominantly due to the high abundance of the Ascomycota taxa Leotiomyces and the Basidiomycota lineage Agaricomycota within the low pH groups. Their abundance decreased dramatically as pH increased and the above ground plant community changed. Pyrosequencing analyses of fungal communities across different land uses has shown that the proportional abundance of the Leotiomyces is relatively consistent across all samples analysed (Lauber *et al.* 2008). However within the study by Lauber *et al.* (2008), there was no difference in soil pH between the study sites. Furthermore the Agaricomycota have been shown to be statistically linked to soils with high C:N ratios, as also seen here within the low pH organic soils. Therefore when the findings of Lauber *et al.* (2008) are combined with the

findings of this study, one may hypothesise that the proportional abundance of the Leotiomyces is controlled to a large degree by soil pH in natural soil ecosystems, but the proportional abundances of other fungal taxa may be influenced by different edaphic properties. If comparisons are drawn between this study and similar studies performed upon bacterial communities the same patterns are evident. Low pH organic soil samples are dominated by a reduced group of bacterial phyla and classes (predominantly acidobacterial lineages) which then causes these soils to exhibit very different community patterns from those of the medium and high pH samples (Chapter 3 - 4). Chao1 richness scores were used as an indication of fungal diversity. Low pH organic soils consistently exhibited lower Chao1 scores than the medium and high pH sample groupings. Similar patterns of increasing Chao1 scores with pH group were seen when bacterial pyrosequencing and fungal pyrosequencing studies were compared (Chapter 4 and 6), suggesting that similar restrictions in diversity are imposed upon both kingdoms, although the mechanisms by which this is achieved may be quite different.

Analysis of fungal biomass across land uses has shown that fungal to bacterial biomass ratios were consistent across all land use types assessed (Lauber *et al.* 2008). Compositional changes were noted between the land use types analysed, but these were linked to edaphic properties rather than the land use type itself. Low C:N, improved soils showed a dominance of the Sordariomycetes, while Agaricomycota taxa dominated Basidiomycota lineages in organic soils with high C:N ratios (Lauber *et al.* 2008). These patterns were consistent with this study, except that C:N ratio was not a significant factor linked to total community variability. Soil pH was found to be more important in determining fungal community structure, along with the above ground plant community, than C:N ratio. Within this study low pH soils had higher C:N ratios than high pH soils, but the only significant variable linked to soil carbon content was loss on ignition. When fungal to bacterial biomass ratios were examined over an artificial pH gradient with minimal differences in other environmental factors, fungal biomass and growth rates were greater within low pH soils than high pH soils (Rousk *et al.* 2010). Rousk *et al.* (2010) showed that although fungal biomass and growth rates altered with increasing soil pH, fungal community composition was reasonably uniform across the pH gradient. This

suggests that although soil pH affects the growth rates and fungal to bacterial biomass ratios, changes in the total community composition between soils cannot be solely attributed to soil pH. However, the system examined contained a single soil type (arable) which underwent long term pH manipulations via lime addition, and thus does not represent soils found naturally in the environment. Furthermore the sample locations were at adjoining positions, which could influence results if soil fungal communities exhibit some form of spatial organisation. Examining the relative effects of multiple environmental variables, at discrete sampling locations, may be key in determining overarching patterns of soil microbial biodiversity as many edaphic and climatic variables interact through complex co-correlating processes (Jenny, 1941). Here, multiple environmental variables are examined concurrently using soils from geographically isolated natural systems and results suggest that several environmental variables are strongly linked to fungal community structure, including soil pH. The co-correlating nature of environmental variables found in natural systems poses problems when trying to partition the proportions of variance in fungal community structure explained by each variable individually (Mitchel *et al.* 2011), but the results provide a holistic view of the most important factors that structure fungal communities. Acidic soil conditions are generally non favourable for optimum intracellular functions as a large number of proteins exhibit denaturation at extremes of pH (Cotter & Hill, 2003). However, many soil fungal isolates have been shown to have a wide tolerance to pH in pure culture studies (Sletten & Skinner, 1948; Hesse & Ruijter, 2002), suggesting that direct interaction with hydrogen ions within acidic environments is not the cause of changes in community composition. However, soil pH affects numerous other edaphic properties including the above ground plant communities (highlighted here as important determinants of fungal community structure) (Partel, 2002), solubility of toxic metals (Pina & Cervantes, 1996), and the size of the labile phosphorus pool. Low labile phosphorus levels have been linked to a high abundance of Sordariomycetes in pasture soils (Lauber *et al.* 2008), and here it can be seen that high proportional abundances of the Sordariomycetes group are seen in high pH soils with theoretically low levels of labile phosphorus (due to high pH conditions), corroborating this finding. Therefore it can be hypothesised that the effect of soil pH is not consistent across different fungal lineages as some, such as the Leotiomycetes, may be directly affected by soil

pH, whereas others such as the Sordariomycetes may be indirectly affected through processes mediated by soil pH. The predominance of the Leotiomyces in low pH conditions suggests a competitive advantage over other fungal taxa within these conditions. Thus, further research is required to determine the exact effect of acidic soil conditions upon fungal community structure, and the mechanisms which cause these differences.

6.4.2 Plant Effects on Fungal Communities

There is a growing body of research that suggests that plant species is a key determinant of soil fungal community structure (Cullings & Hanely, 2010; Mitchel *et al.* 2011; Moora *et al.* 2011). Within this study, the above ground plant community composition was seen to be the strongest predictor of fungal community structure, which corroborates findings presented in Chapter 5 using a landscape scale t-RFLP approach. A large proportion of land plants form symbiotic relationships with fungal organisms (Malloch *et al.* 1980), but the level of plant host specificity of fungal symbioses is in question as individual plant species have multiple potential fungal symbionts within temperate soil ecosystems, whereas plant fungal symbiont diversity is substantially lower in tropical soil ecosystems (Comas *et al.* 2010). Further work may include similar studies examining environmental influences upon soil fungal populations within temperate and tropical soil ecosystems to determine if fungal populations follow similar patterns of biodiversity relating to latitudinal gradients to those seen in macro organisms (MacArthur, 1972; Diaz *et al.* 1998). As a large number of fungal lineages have direct associations with plant roots, the above ground plant community will influence these organisms more than fungal lineages which do not directly interact with plants. Separately analysing patterns of plant and mycorrhizal fungal biodiversity from lineages without known mycorrhizal associations would highlight differences in the relative effects of edaphic properties upon total fungal community structure. Furthermore it is well documented that specific plant lineages have different tolerances to soil pH (Partel, 2002) as well as other variables identified here as important determinants of fungal community structure (for example soil moisture). This highlights the difficulty of determining the exact cause of fungal community shifts across natural environmental gradients, but provides a useful insight into

the main factors which must now be examined in more detail at the landscape scale.

Potentially the proportional abundance and distribution of different fungal lineages may be governed by different environmental variables. Further large scale assessments of fungal community composition and structure, using taxonomically informative methods, may be required to accurately determine the relative effects of numerous co-correlating environmental variables upon fungal community structure and individual fungal lineages.

6.4.3 ITS and 18S rRNA Gene Comparisons

Phylum level patterns in fungal biodiversity exhibited distinct differences between the ITS sequencing approaches presented in Chapter 5 and the 18S rRNA sequencing approach presented here. Ascomycota lineages dominated samples from low and high pH groups, in contrast to the dominance of Basidiomycota lineages within the low pH environments seen in Chapter 5. However, class level patterns were consistent between the two studies. Within the low pH organic soils, the Leotiomycetes were the dominant Ascomycota lineage detected, and the proportional abundance of this class decreased across the pH groups. A dominance of the sister taxon Sordariomycetes was seen within the medium and high pH samples, as was an increase in the Basidiomycota lineage Tremellomycetes. Agaricomycota were the dominant Basidiomycota class within the low pH samples, while Chytridiomycota lineages were most prevalent within the high pH samples, also corroborating findings presented in Chapter 5.

6.4.4 Global Differences in Dominant Fungal Lineages

Although few assessments of patterns in fungal biodiversity have been attempted using pyrosequencing (Lentendu *et al.* 2011), differences between the most dominant fungal phylum in soils have been shown between soil ecosystems. Pyrosequencing studies of temperate soil ecosystems concur that a dominance of Dikarya exists (Lim *et al.* 2010; Buée *et al.* 2009). However, Arctic tundra soils have been shown to have a dominance of Zygomycota lineages (Chu *et al.* 2011), suggesting that global patterns of fungal biodiversity are different from landscape scale assessments. In contrast, within this study

Ascomycota dominated low and high pH samples while Basidiomycota dominated medium pH samples, showing that edaphic variability has a significant effect on the dominant fungal group. Differences in the gene sequenced between this study and that of Buée *et al.* (2009) potentially explain this result, as does the fact that Buée *et al.* (2009) examined only high C:N forest soils. Basidiomycota lineages have been shown to be in high abundance within these soil types in other studies (Lauber *et al.* 2008). Earlier versions of pyrosequencing chemistries were used in the study by Lim *et al.* (2010), which produced fewer reads with much shorter read lengths than those used here, potentially affecting accurate taxonomic assignment. Accuracy in taxonomic assignments using the 18S rRNA gene has been shown to increase dramatically as read length increases (Liu *et al.* 2011), supporting this hypothesis.

6.4.5 Fungal OTU Sharing

Overlap of fungal taxa between sites is not often studied. However, an 18S rRNA clone library analysis of soil fungal populations across 3 sites in the USA (forest, desert, and tall grass prairie) using between 200 and 400 sequences per sample found no overlap in fungal OTUs at 97 % OTU similarity (Fierer *et al.* 2007). Here, using pyrosequencing analysis of the same gene but examining 15 sites and returning in the region of 2,000 – 5,000 sequences per sample, OTU sharing between sites was present but minimal when analysed at the same level of OTU sequence similarity. The radial trees show that most OTUs were restricted to a single pH grouping of samples irrespective of lineage; however, a large proportion of Ascomycota OTUs were unique to the low pH samples while Basidiomycota and Chytridiomycota OTUs were predominantly restricted to the medium and high pH samples. When OTUs were clustered at lower levels of sequence similarity (90 %) again a high level of endemism was seen. OTUs shared between the medium and high pH groups were largely absent from the low pH group and vice versa. OTUs that were shared across all samples of one pH group were comparatively rare, supporting findings by Fierer *et al.* (2007) and those presented in Chapter 5 suggesting that fungal evenness is low and there is a high level of OTU variability between samples when analysed at the landscape scale. However, when taxa abundance was summarised from OTUs with assigned identifications at the class level, distinct

patterns emerged with reference to pH group. Indicator taxa analysis highlighted a number of fungal lineages which were predominantly found within one pH group but were largely absent from the others, indicating that edaphic properties influence patterns in fungal biodiversity at the phylum and class level.

6.4.6 Explanations for Changes in Fungal Composition Across Environmental Gradients

A general understanding of the properties of each indicator taxon goes some way to providing hypotheses as to why each was found predominantly within one pH group and not in others. All taxa defined here as indicators of low pH soils belonged to the Ascomycota which, in general, were at a higher proportional abundance than any other fungal phylum within these samples. More specifically, each of these indicator taxa belonged to classes found within the subdivision Pezizomycotina. Many members of this class have been shown to be linked with particular plant species, in the form of root symbionts, endophytes, pathogens, ectomycorrhizal parasites, or wood rot fungi (Wang *et al.* 2006). As there is a distinct plant/fungi relationship within this class, it is unsurprising that plant biogeography has an effect on the distribution of these taxa. However, as soil pH is a major determinant of above ground plant communities (Partel, 2002), both these factors are important but the exact mechanisms causing the changes in fungal community structure are unclear.

Providing theories to explain the differences in the proportional abundance of the Laboulbeniomyces between the pH groupings is troublesome. They are external parasites of insects and other arthropods (Henk *et al.* 2003) but many have life stages spent in soil. Soil pH may have a direct effect upon the proliferation of these taxa during a soil-borne stage of the life cycle, or indirect effects by altering the local distribution of soil insects. Meta-analysis of the distribution of known laboulbeniomyces host insects may provide an explanation for the higher abundance of these parasitic fungi in low pH soils. The Arthoniomyces are predominantly rock inhabiting fungi, sometimes symbiotically linked with lichens. Oddly these organisms are reportedly found predominantly in tropical and subtropical regions (Spatafora *et al.* 2006). Ruibal *et al.* (2009) state that most rock inhabiting fungi are known for tolerance to harsh conditions, potentially explaining their abundance in low pH soils (Ruibal

et al. 2009). However, the generally understudied nature of rock inhabiting fungi hinders a full understanding of the distribution of these organisms. Fungi in the class Glomeromycetes, an indicator taxon of the medium pH soils, are generally regarded as obligate biotrophs dependent on symbiotic relationships with land plants (Schüßler *et al.* 2001), further supporting the evidence that the above ground plant communities have an important role in determining patterns of fungal biodiversity. Other fungal taxa identified as important descriptors of soil pH here are not wholly dependent upon plant based symbiosis. For example, sporocarpic members of the Endogonaceae, seen here as indicator taxa of the medium pH soils, have been shown to be adapted for dispersal by rodents (Janos *et al.* 1995). Spores from members of the Glomeromycota are, in general, comparatively large and thus aerial dissemination is less efficient than vector mediated dispersal. The higher abundance in medium pH soils may then, in part, be related to limited dispersal mechanisms as the distribution of vectors may also be restricted by soil pH (possibly by food availability to rodents from agricultural practices).

Fungi in the class Tremellomycetes, an indicator taxon of the medium pH samples, are known to be saprophytic and parasitic towards animals and fungi (Millanes *et al.* 2011). As little research has been directed towards members of this lineage using molecular methods (Matheny, 2006) they require further analysis as they have been identified here as important taxa within soils of neutral pH.

Neocallimastigomycota are obligate anaerobes found within the digestive system of herbivores (Liggenstoffer *et al.* 2010). Neutral pH soils are commonly used as areas for grazing pasture due to the increased plant productivity achieved through a more efficient uptake of fertilisers (Vickery *et al.* 2001). Therefore the abundance of organisms belonging to this lineage may be directly linked to the abundance and distribution of grazing animals and shows that soils are a potential reservoir for these organisms. Neocallimastigomycota play essential roles in the degradation of fibre within the digestive systems of herbivorous animals (Liggenstoffer *et al.* 2010), achieved using polysaccharide degrading enzymes such as cellulases, hemicellulases and xylanases, which hydrolyse even the most recalcitrant plant polymers (Akin & Borneman, 1989). If, as suggested here, these organisms reside within the soil ecosystem in a

viable form, environmental reservoirs may exist which provide an essential transmission route between individual herbivorous livestock.

The Chytridiomycetes were the only indicator taxon of high pH soils, which corroborates findings presented in Chapter 5 highlighting their abundance within these soil types. The Chytridiomycetes are flagellate fungi and have recently been redefined to exclude the Neocallimastigomycota. They are globally distributed in soils, freshwater, and saline estuaries (James *et al.* 2006). Very few OTUs belonging to the Chytridiomycetes were found in the low pH samples. Of the OTUs that were detected in the low pH samples, all belonged to the Spizellomycetales or Chytridiales orders. This in itself suggests a direct inhibition of growth in acidic organic soils, or a reduction in abundance due to being out competed for limited resources by other dominant members of the fungal biosphere, for example the Leotiomycetes. Chytridiomycetes are generally saprophytic and are thus important in nutrient cycling as they have the ability to break down keratin and chitin (James *et al.* 2006). Several chytridiomycetes are known to have parasitic relationships with invertebrates, algae, and plant roots. However, due to the incomplete taxonomic classification of sequences within this study, it is difficult to derive patterns or provide hypotheses explaining the distribution of this lineage. In general, OTUs classified as belonging to the Chytridiomycota were detected in proportionally higher abundances within the medium and high pH soil samples with a distinct preference for the latter.

6.4.7 Conclusions

Here it has been shown that a biogeographic effect in relation to soil fungal communities does occur, suggesting that external variables affect the community composition of fungal populations, namely above ground plant communities, soil pH, loss on ignition, and soil moisture. It may be the case that some fungal taxa respond to plant biogeography to a greater extent than others, whereas those with little direct interaction with plants may be structured to a greater degree by edaphic conditions (Mitchel *et al.* 2011). Therefore, if the presence and abundance of only a proportion of fungal lineages are determined by plant fungal interactions, further research is required to understand the mechanisms by which fungal lineages not associated with plants (e.g. rock

inhabiting fungi, or parasitic/mutualistic fungi of animals or other fungi) are structured. Here it is proposed that, like soil bacteria and archaea, an important factor may be soil pH. It may also be that plants do directly determine the fungal community present within a site via the input of metabolites such as root exudate compounds (Broekling *et al.* 2007); however, the bioavailability of these compounds to specific fungal taxa may be mediated by edaphic variables such as soil pH. Conversely, a large number of research papers are directed towards the idea that fungal communities directly influence plant diversity and community structure (reviewed by van der Heijden *et al.* 2008). It therefore seems a circular argument akin to the chicken and egg paradox: do plant species govern fungal community structure or vice versa? A more detailed analysis regarding the effect of the above ground plant community structure upon the total below ground community structure is needed along with measurements of edaphic and climatic properties.

6.4.8 Future Work

Along with proposed future studies documented above, to fully understand biogeographic patterns within the soil microbiota, the effect of spatial autocorrelation must also be examined. Fungal and bacterial organisms exhibit very different physical and life cycle properties and thus the distance between sites examined may affect the biogeographical patterns observed in landscape and global scale studies into microbial biodiversity. Dispersal mechanisms and the individual size of members of each kingdom are very different, suggesting that dispersal rates and ability (key determinants of spatial auto correlation) may show distinct patterns in relation to spatial distribution.

Chapter 7: Investigating Spatial Autocorrelation in Soil Bacteria and Fungi

7.1 Introduction

7.1.1 Background

The principal aim of studies examining the biogeography of organisms is to define the distribution of species in terms of environmental variables presumed to be the operative constraints on specific taxa. However, geographic separation is also thought to affect the community structure of particular organisms exclusive of environmental constraints. Tobler's first law of geography simply states "Everything is related to everything else, but nearby objects are more related than distant objects" (Miller, 2004). The theory of spatial autocorrelation is inextricably linked with this statement and, when applied in a biological context, it is thought to arise through a number of processes not determined by environmental variables, but by geographic separation alone. Spatial autocorrelation is defined as the correlation amongst values of a single variable strictly attributable to the proximity of those values in geographic space (Griffith, 2003). Positive spatial autocorrelation is where values of random variables, at pairs of locations certain distances apart, are more similar than expected for randomly associated pairs of observations, with negative spatial autocorrelation being the opposite (Legendre, 1993). Spatial autocorrelation within ecological systems may be due to limitations to dispersal, predation, social aggregation, food availability, parasitism, evolution and adaptation, migration and extinction, or competition (Legendre & Fortin, 1989). For investigations of relationships between the environment and community structure, spatial autocorrelation is often viewed as a confounding factor. However, spatial autocorrelation itself is information bearing as it reveals associations between geographic entities (Miller, 2004). To expand, the detection and quantification of spatial autocorrelation in organisms is of particular interest as it allows a greater understanding of the mechanisms by which complex communities are structured (Ettema & Wardle, 2002). Furthermore, as the spatial distributions of soil microbes and microbially mediated functions have high intrinsic variability, understanding the spatial ordering of soil microbial taxa helps reveal the true effects of external processes on soil microbial ecology, such as disturbance, management practices or

climate change (Nunan *et al.* 2002). At the local scale, spatial heterogeneity is seen to be functional through biotic processes such as those discussed previously (such as predator-prey interactions and food availability). Therefore, ecosystems without spatial structuring would be unlikely to function (Legendre, 1993).

Many different organisms have been shown to exhibit some form of spatial autocorrelation, for example plants (Dirnbock & Dullinger, 2004), birds (Raymond *et al.* 2006), mammals (Peakall *et al.* 2003), and insects (Johnson & Worobec, 1988). However, few studies have explicitly examined spatial autocorrelation within temperate soil microbial populations, especially at large geographic scales (Wirth, 1999). In previous chapters soil bacterial and fungal communities have been shown to have links with a number of co-correlating environmental variables, and thus they are hypothesised to be structured in a similar fashion to macro organisms. To further this comparison and adapt ecological theories derived from observations of larger organisms onto these complex populations, the effect of geographic separation upon the structuring of microbial communities should be examined. If spatial autocorrelation is observed in soil microbial communities, it is likely that some or all of the aforementioned mechanisms inducing spatial autocorrelation may also influence the overall biogeography of these organisms at the landscape scale.

Using bacterial targeted pyrosequencing to examine 85 soils located in a high alpine environment up to 2 km apart, King *et al.* (2010) showed that significant spatial auto correlation occurred up to 240 m. Furthermore, individual groups of soil bacteria were seen to be organised in a spatial manner. Using a further 6 soil samples from 4 separate mountain habitats across the globe, Sanger sequencing produced a global model of bacterial spatial autocorrelation (King *et al.* 2010). Small scale spatial variation in soil bacterial and fungal populations has also been exhibited in other instances, but these studies explicitly examined single habitats: forests (Lilleskov *et al.* 2004) or arable soils (Nunan *et al.* 2002).

A problem with directly assessing spatial autocorrelation in any given set of organisms is that environmental variability is often co-correlated with spatial variability. For example, if Tobler's first law of geography is applied to the environment, as was originally intended, then habitats located close together

also exhibit a higher degree of similarity than those further apart. As environmental variability is directly related to spatial variability (Legendre, 1993), and community dissimilarity potentially linked to environmental dissimilarity, a spatial effect upon community dissimilarity may be incorrectly reported if the effect of environmental dissimilarity is not accounted for. A number of statistical tests have been proposed to remove the effect of the environment or space upon the structuring of communities in order to examine one or the other exclusively. These include the popular Mantel test (Mantel, 1967; Legendre, 2000), calculation of semivariogram models, and principal coordinates analysis of neighbourhood matrices (Dray *et al.* 2006).

The Mantel test has become the workhorse of spatial ecology and quite simply quantifies the level of correlation between two or more dissimilarity or similarity matrices by calculating a Pearson's product moment correlation coefficient. In order to define a spatial or ecological effect upon a community, dissimilarity matrices relating to taxa composition, environmental variability, and spatial separation can all be generated using appropriate dissimilarity or similarity measures, then correlations can be computed between each. In order to examine the effect of one of these variables exclusively on any of the other two, the third must be removed or "partialled" out of the calculation. This can be achieved using a partial Mantel test, whereby the residual variation of the correlations between two of the matrices is correlated with the third matrix of interest. In the case of spatial autocorrelation, first, the correlation between the community based dissimilarity matrix and the environmental dissimilarity matrix is calculated, and then the residual variation in the former matrix is correlated with the spatial dissimilarity matrix. Permutation tests may be carried out to discover whether the correlation coefficients (reported as a Mantel Statistic) are significantly different from 0 by permuting the number of columns and rows within each of the dissimilarity matrix numerous times (Oksanen, 2011). If a significant positive correlation coefficient is observed it can be said that, as geographic dissimilarity (i.e. the space between samples) increases so does community dissimilarity (i.e. the community composition becomes more different). However, overall partial Mantel tests only take into consideration the average correlation across the entire data set (Urban, 2003). In order to calculate Mantel statistics for discrete distance classes to identify at which

distances spatial autocorrelation occurs, a multivariate correlogram can be constructed from a Mantel or partial Mantel test. Here the correlation between two dissimilarity matrices (removing the effect of a third if desired) is calculated for discrete distance classes by splitting up the data points based upon their respective geographic separations. Distance matrices representing community dissimilarity, environmental dissimilarity, and geographic dissimilarity are then correlated in the desired fashion to determine at which particular distances spatial autocorrelation occurs (if at any).

7.1.2 Aims

Large scale bacterial and fungal targeted t-RFLP studies were undertaken to examine the effects of spatial autocorrelation upon the structuring of soil microbial communities in these two kingdoms. Also, DGGE was used, as a molecular community profiling method with a higher taxonomic resolution than t-RFLP, to explore spatial autocorrelation within the bacterial kingdom over two distinct soil ecosystems with vastly different bacterial populations and levels of community variability.

7.2 Materials and Methods

7.2.1 Sample Selection

T-RFLP profiles generated in Chapters 3 and 5 were used to examine spatial autocorrelation in bacterial and fungal populations. For bacterial targeted DGGE analysis two subsamples representing distinct habitat types, shown to harbour very different bacterial communities, were selected from a subset of the Countryside Survey 2007 samples (Figure 49). Each subsample represented a singular habitat type dominated by either *Calluna* or *Lolium*. The *Calluna* dominated subsample comprised 32 15 cm by 5 cm soil cores taken from acidic bog regions with high organic matter and soil moisture contents. The 32 *Lolium* dominated samples were taken from typical improved soils defined as neutral grassland with high mineral content and comparatively low levels of organic matter. Chapter 3 shows how these soil types harbour distinctly different bacterial communities, and that the variability of bacterial communities within each of these habitats is very different: low pH bogs have higher bacterial community variability than neutral and high pH grasslands.

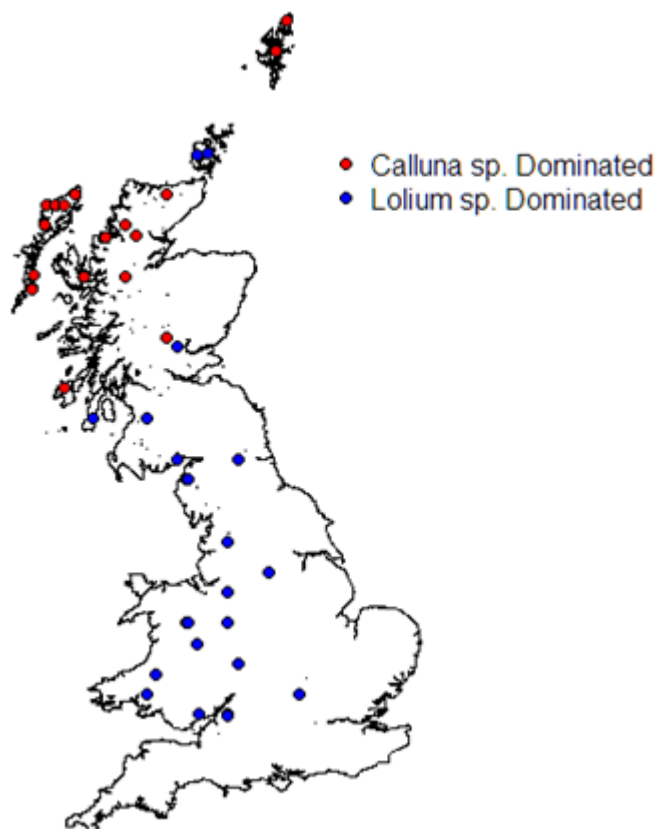


Figure 49: The 32 *Calluna* dominated soil samples (red) and 32 *Lolium* dominated samples (blue) used for DGGE analysis of bacterial populations for an in depth analysis of spatial autocorrelation of two distinct soil biomes shown to harbour very different bacterial communities.

7.2.2 DNA Extraction

Nucleic acid extractions were carried out as detailed in Chapter 2.2.2.

7.2.3 T-RFLP Analysis

T-RFLP profiles were generated as described in Chapter 3.2.5 for bacterial profiles and Chapter 5.2.4 for fungal profiles.

7.2.4 DGGE

Approximately 50 ng of nucleic acid extract was used as a template to amplify the 200 bp length of DNA spanning the V3 region of the 16S rRNA gene using the primers 338GC (5'-CGC CCG CCG CGC GCG GCG GGC GGG GCG GGG GCA CGG GGG G -3') (Muyzer *et al.* 1993) and 519r (5'- GTA TTA CCG CGG CTG CTG - 3') (Lane *et al.* 1991) (Sigma Genosys, Poole, UK). Reactions were carried out with 1 pmol μl^{-1} of each primer, 2.5 U of Taq polymerase (Sigma, Poole, United Kingdom), deoxynucleotides at a concentration of 200 μM each, 1 mg ml^{-1} BSA, 5 μl Taq buffer containing 1.5 mM Mg^{2+} with 41.9 μl PCR H_2O (Sigma Genosys, Poole, UK). The following reaction conditions were used: an

initial denaturation step of 95 °C for 1 min 30 s, then 35 cycles of 95 °C for 60 s, 60 °C for 45 s and 72 °C for 90 s, and a final elongation step of 72 °C for 30 min.

Approximately 1 µg of PCR amplified DNA product was loaded into a randomly allocated well of a 10 % w/v polyacrylamide with a urea based denaturing gradient ranging from 30 % to 60 %. A gradient was formed by mixing 30 % denaturant (12.6 g urea, 12 ml formamide, 25 ml 40 % bis-acrylamide solution (37.5:1 acrylamide: bisacrylamide), 2 ml 50 x Tris-Acetate-EDTA (TAE) in 100 ml ddH₂O) with 60 % denaturant (25.2 g urea, 24 ml formamide, 25 ml acrylamide, 2 ml 50x TAE in 100 ml ddH₂O) using a peristaltic pump (Gilson, WI, USA) and gradient former (BioRad, CA, USA). Polymerisation of acrylamide was achieved using 12 µl of tetramethylethylenediamine (TEMED) with 0.1 g of ammonium persulphate (APS) dissolved in 500 µl dd H₂O. Electrophoresis was carried out in the Dcode system (Bio-Rad, CA, USA) at a constant 100 V for 18 h at 60 °C. Gels were then stained using 2 µl 10,000 x SYBR Gold (Invitrogen, Grand Island, USA) in 20 ml TAE for 30 min. Gel images were digitised using a Versadoc imaging system (Bio-Rad, CA, USA) and bacterial abundance was examined based upon relative band intensity using computer aided manual selection of DGGE bands in the computer software Phoretix 1D (Phoretix International, Newcastle upon Tyne, United Kingdom).

7.2.5 Statistical Analyses

All statistical analyses were performed in R using the packages *ecodist*, *vegan*, *cluster* and *MASS*. Taxa dissimilarity matrices were constructed using a Bray-Curtis dissimilarity measure; dissimilarity matrices representing geographic separation of samples were constructed using Euclidean distance measure; and environmental dissimilarity matrices were computed using a Gower dissimilarity index to include categorical variables such as habitat type. All Mantel statistics were computed using the Pearson's product moment correlation co-efficient as suggested in the *Vegan* documentation (Oksanen, 2011). Tests of significance by permutations ($n = 999$) of N rows and columns of the dissimilarity matrices were automatically included in calculations. Mantel and partial Mantel correlograms were constructed using the *ecodist* package with 999

permutations. Mantel statistics with a significant difference from 0 ($p < 0.01$) are displayed as solid points within the plots.

7.2.6 Overview of the Mantel Statistic

As previously mentioned, the Mantel statistic aims to quantify the correlation between two distance matrices while, in the case of the partial Mantel test, controlling for the effects of a third. As the distance matrices are symmetrical (i.e. the distance from point A to point B is the same as from point B to point A) each distance is not independent of every other distance. Therefore direct significance tests between matrix dissimilarities cannot be performed as changing the position of one distance alters the positions over all other distances. The Mantel test circumvents this problem by correlating the distance matrices while randomizing the columns and rows of one matrix and keeping the other constant multiple times. A significant deviation from no correlation (i.e. a correlation coefficient of 0) can then be identified if the primary correlation coefficient is greater than those generated from the randomised data set at a level of significance defined by the user. The Mantel statistic itself is generated as a normalized Pearson's product moment correlation coefficient where variables of different units are scaled so correct comparisons can be made. This is calculated using the following equation: -

$$r = \frac{1}{(n-1)} \sum_{i=1}^n \sum_{j=1}^n \frac{(x_{ij} - \bar{x})}{S_x} \cdot \frac{(y_{ij} - \bar{y})}{S_y}$$

Where x and y are variables measured at locations i and j , and n is the number of elements within each of the distance matrices. Scaling of variables is achieved by dividing each variable at location i and j from matrix x and matrix y by its respective standard deviation (S). A Mantel statistic (r) within the bounds of a normal correlation coefficient (between -1 and 1) is provided as a result of the first part of the calculation, i.e. by multiplying the summation of each matrix calculation by $1/(n-1)$.

7.3 Results

7.3.1 Analysis of Bacterial Spatial Autocorrelation using t-RFLP

A significant overall positive correlation between bacterial community dissimilarity and geographic distance was observed using all soil samples

(overall Mantel statistic = 0.232, $p < 0.001$). Positive Mantel statistic values were observed up to approximately 400 km. At distances greater than 400 km significant negative Mantel statistic values were seen (Figure 50). This suggests that bacterial community composition is structured differently incrementally up to a certain geographical distance and becomes increasingly more dissimilar at intermediate distance classes. Furthermore, Mantel statistics for the largest distance classes are less negative than those at intermediate distance classes, suggesting that soil bacterial community dissimilarity is greatest over intermediate distance classes.

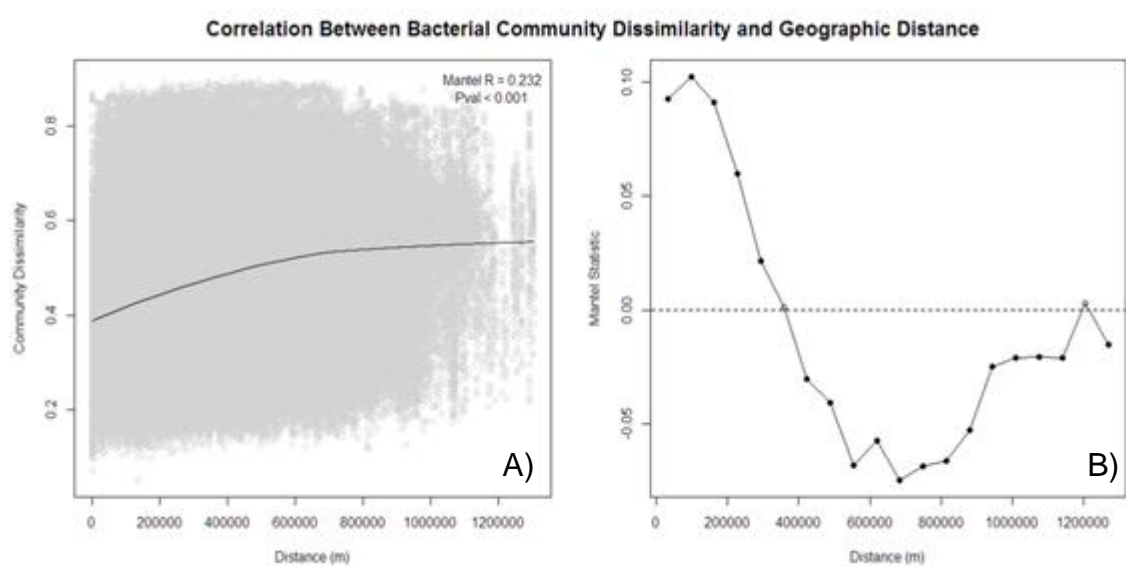


Figure 50: A) Mantel test showing a significant positive correlation between bacterial community dissimilarity and increasing distance (Mantel statistic = 0.232, $p < 0.001$). A locally weighted polynomial regression line is fitted to show this correlation. B) Mantel correlogram showing that positive spatial autocorrelation predominantly occurs at smaller scales up to approximately 400 km. Filled circles indicate values that are significantly different from zero; open circles indicate values that are not significantly different from zero.

However, this initial analysis does not take environmental correlations with geographic distance into consideration. Environmental dissimilarity was also shown to increase with geographic distance (Figure 51) (overall Mantel statistic = 0.33, $p < 0.001$). When environmental variability was partitioned into discrete distance classes, again significant positive Mantel statistic values were observed at distances up to 40 km, and negative Mantel statistic values thereafter. Therefore environmental dissimilarity is greatest at intermediate distances, as is the case with bacterial community dissimilarity, but at the largest distance classes the Mantel statistic becomes less negative. This suggests that environments situated at large distances apart are more different than those located within close proximity to each other, but are more similar to each other than those found at intermediate distances. Bacterial community

dissimilarity was also seen to have significant strong positive correlations with environmental dissimilarity (total Mantel statistic = 0.625, $p < 0.001$), showing that as the environment becomes more different, so does the community composition of the bacterial populations residing within (Figure 52). When this correlation was split into discrete distance classes, the Mantel statistic was positive up to an environmental dissimilarity value of 0.2. From then on, significant negative values were observed, but values for the largest dissimilarity classes were less negative, again suggesting that bacterial community dissimilarity is greatest at intermediate classes of environmental dissimilarity. This is counter intuitive as one would expect communities residing within the most different habitats to be the most different in composition; however, this may be due to a limitation in bacterial community variability. To elaborate, the largest differences between bacterial communities detected by t-RFLP in Chapter 2 were due to changes in the abundance of a few TRFs. Therefore the bacterial community structure as defined by t-RFLP can only become so different with changing environmental conditions.

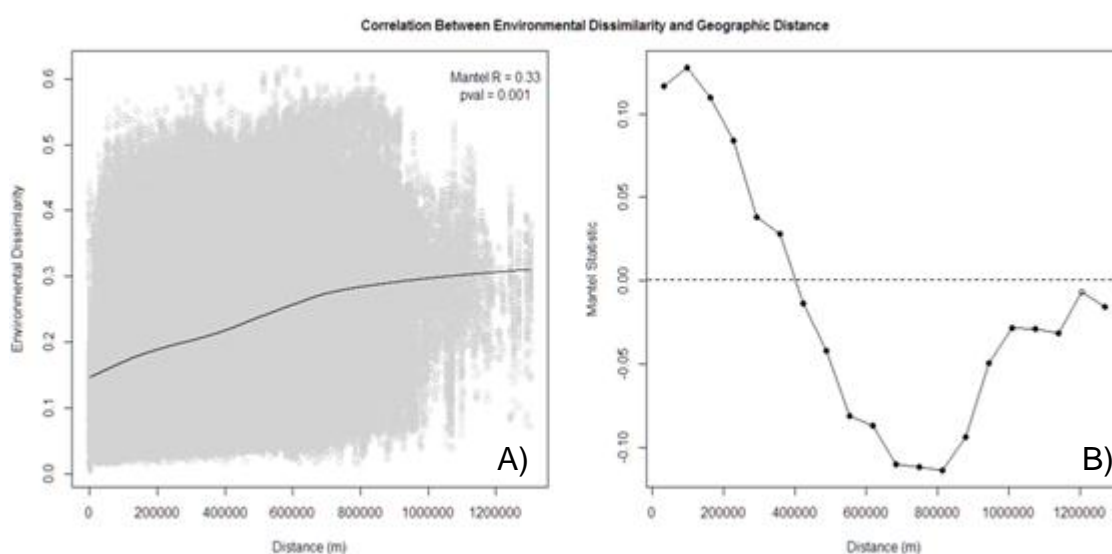


Figure 51: A) Significant correlation between environmental dissimilarity and geographic distance showing that as distance between sites increases so does environmental dissimilarity (Mantel $R = 0.33$, $p < 0.001$). B) Mantel correlogram showing that strongly positive Mantel statistics are seen up to approximately 400 km. Negative Mantel statistics at intermediate geographic distances suggest that environmental dissimilarity is greatest over intermediate distances. Filled circles indicate values that are significantly different from zero; open circles indicate values that are not significantly different from zero.

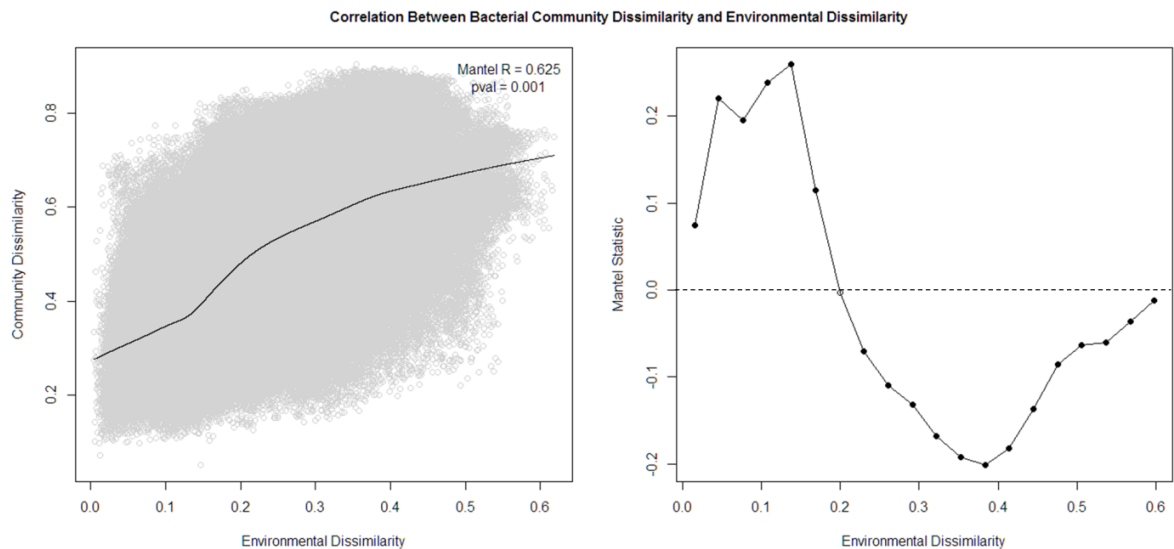


Figure 52: A) Bacterial community dissimilarity increases with environmental dissimilarity (Mantel $R = 0.625$, $p < 0.001$). A locally weighted polynomial regression shows that this effect is most obvious over the first increments of environmental dissimilarity. B) Strongly positive Mantel statistics are seen up to an environmental dissimilarity value of 0.2 and negative values thereafter. Filled circles indicate values that are significantly different from zero; open circles indicate values that are not significantly different from zero.

To define the relationship between bacterial community dissimilarity and geographic distance alone, discounting the effect of environmental dissimilarity, partial Mantel tests and correlograms were implemented. Here the correlations between environmental variability and geographic distance are removed leaving the effect of pure space upon bacterial community dissimilarity. Figure 53 shows that over short geographic distances (up to approximately 250 km), positive spatial autocorrelation occurs after the removal of measured environmental variability, and from 250 km to 600 km negative spatial autocorrelation occurs. This provides further evidence to suggest that bacterial community dissimilarity is greatest at intermediate distance classes.

Partial Mantel Correlogram Describing the Correlation Between Bacterial Community Dissimilarity and Geographic Distance While Controlling for Environmental Variability

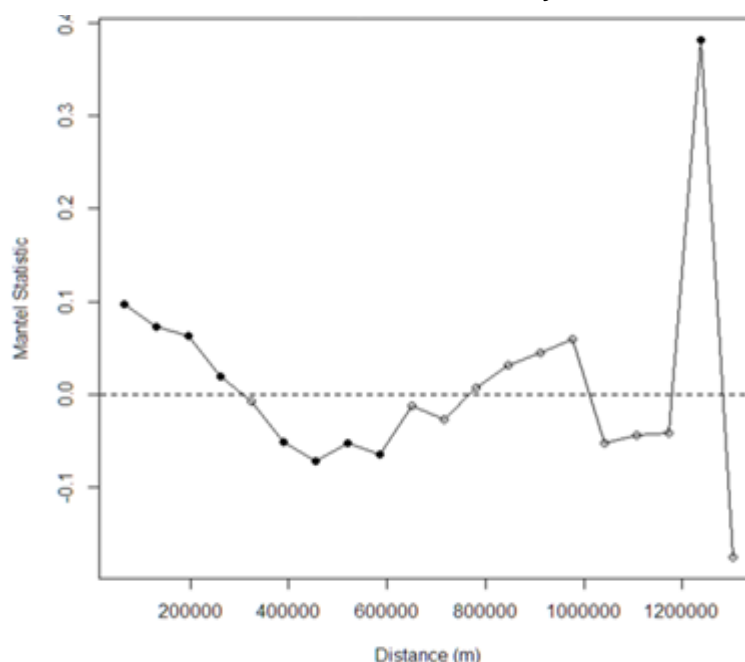


Figure 53: Partial Mantel correlogram showing that, within soil bacterial communities, positive spatial autocorrelation occurs at distances up to approximately 250 km. At greater distances, strong positive correlations were also observed. Filled circles indicate values that are significantly different from zero; open circles indicate values that are not significantly different from zero.

When bacterial communities in samples 1200 km apart were analysed, significant strongly positive Mantel statistics were observed. This seems to show that communities from samples situated at distances very far apart exhibit the strongest positive spatial autocorrelation. However, the large positive Mantel statistic at this distance class may be a result of a reduced number of samples separated by large distances. Although permutation tests show this finding to be significant, results should be interpreted with caution.

7.3.2 Bacterial Spatial Analysis of Two Distinct Soil Environments using DGGE

Significant positive spatial autocorrelation was observed at small and large distance classes using the bacterial targeted t-RFLP profiles. To analyse this phenomenon in more detail and at a higher taxonomic resolution, DGGE analysis was performed on two subsets of geographically dispersed soil samples each with distinct soil characteristics (subset n = 32). Acidic organic soils dominated by *Calluna* and neutral improved grassland soils dominated by *Lolium* were selected as a subset of the samples used above. These were

defined as representative soils shown to harbour distinct and contrasting bacterial communities (Chapters 2 and 4). Simply by examining the DGGE gels by eye (Appendix S: A - B) the low pH *Calluna* dominated soil bacterial profiles were seen to be much more variable than the *Lolium* dominated soil bacterial profiles. Dissimilarity matrices were calculated for each subsample and the mean variability was displayed in the form of boxplots (Figure 54). A significant difference in mean variability was seen between the two subsamples (t-test $p < 0.001$), supporting the theory stated in Chapter 3 that low pH communities have a higher within pH group variability (beta diversity) than medium and high pH samples.

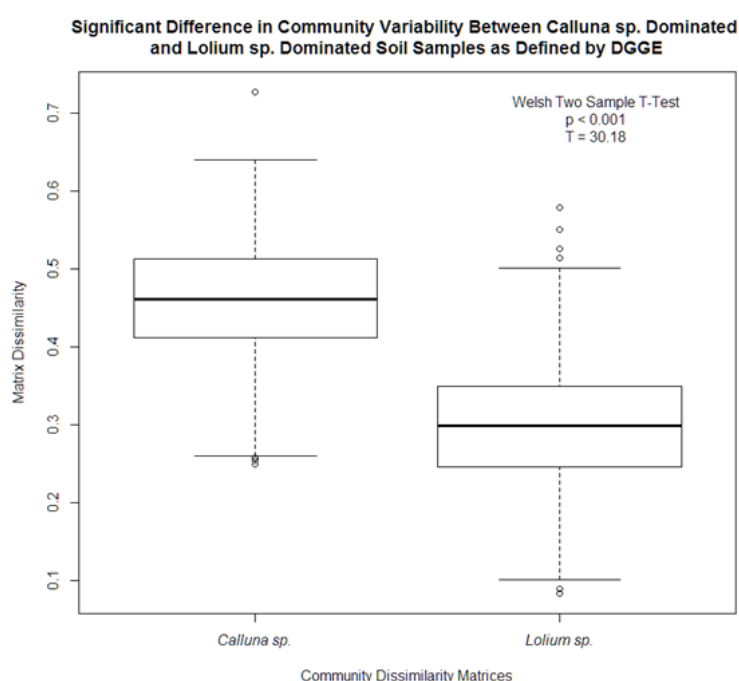


Figure 54: Boxplot showing mean pairwise dissimilarity (Bray-Curtis) for the *Calluna* and *Lolium* dominated subsamples. *Calluna* dominated samples from acidic bog regions show a statistically significant, higher mean pairwise bacterial community dissimilarity than the *Lolium* dominated subsamples from mineral soils supporting neutral grassland ecosystems (t-test $p < 0.001$).

To compare spatial autocorrelation between the two sites and via the use of different community profiling methods, Mantel tests were carried out on the site by species matrix for each DGGE gel and for the t-RFLP profiles generated in Chapter 3 for the same samples (Table 12).

	Matrix Correlation					
	Taxa/Environment		Taxa/Distance		(Taxa/Distance) - Environment	
Site Type	t-RFLP	DGGE	t-RFLP	DGGE	t-RFLP	DGGE
Calluna	R = 0.13 p < 0.01	R = 0.11 p < 0.01	R = 0.18 p < 0.01	R = 0.24 p < 0.01	R = 0.15 p < 0.01	R = 0.23 p < 0.01
Lolium	R = 0.28 p < 0.01	R = 0 p > 0.1	R = 0.05 p > 0.1	R = 0.04 p > 0.1	R = 0.04 p > 0.1	R = 0.04 p > 0.1

Table 12: Mantel R statistics describing the correlation between soil bacterial taxa and environmental dissimilarity, and soil bacterial taxa and geographic separation of two contrasting soil bacterial communities from low pH organic soils (dominated by *Calluna*) and higher pH improved soil (dominated by *Lolium*) using two different molecular profiling methods with different taxonomic resolutions, namely t-RFLP and DGGE. The final correlation column represents a partial Mantel test where the effect of environmental dissimilarity is removed from the taxa distance relationship to highlight the effect of geographic separation alone upon bacterial community structure.

The species-environment relationship was stronger in the t-RFLP assessment than the DGGE assessment for both habitats (*Calluna* dominated or *Lolium* dominated). Even though small environmental relationships were detected using the DGGE method for the *Calluna* dominated samples, there was no significant environmental relationship when the *Lolium* dominated samples were analysed by DGGE. This may be because t-RFLP has a lower taxonomic resolution and thus examines overarching trends in bacterial community structure in relation to environmental variability, whereas DGGE examines community structure closer to the species level and thus may be more sensitive to local variations in environmental characteristics. The low pH *Calluna* dominated samples are relatively variable both in community structure (Figure 54) and environment (Chapter 3) whereas the *Lolium* samples are comparatively uniform across all samples in bacterial community structure.

The species-distance relationship was only significant for the *Calluna* dominated samples using either the t-RFLP approach or the DGGE approach, and the spatial relationships were stronger when DGGE was used than with t-RFLP. No significant spatial relationships were detected with the *Lolium* dominated samples using either method. Again this is likely to be due to the increased taxonomic resolution of DGGE compared with t-RFLP and the fact that the *Calluna* dominated samples are more variable than the *Lolium* dominated samples: if there is little variability in community structure across all samples, samples taken at large distances are going to be the same as

samples taken across small geographic distances, as is the case with the *Lolium* dominated samples. However, the *Calluna* dominated samples have a high intrinsic variability, and that variability significantly increases as geographic separation increases.

Partial Mantel correlograms show that, within the *Calluna* dominated samples, significant spatial autocorrelation was observed using t-RFLP and DGGE. The relationship was stronger when DGGE analysis was used instead of t-RFLP. No significant spatial autocorrelation was seen in the *Lolium* dominated samples. Therefore, even though environmental variability is generally greater in low pH soils, after accounting for that environmental variability, positive spatial autocorrelation is still observed.

7.3.3 Analysis of Fungal Spatial Autocorrelation using t-RFLP

A significant positive correlation between fungal community dissimilarity and geographic distance was observed (Mantel statistic = 0.103, $p < 0.001$). Mantel correlograms show that Mantel statistics are positive at small and large distance classes, but small negative Mantel statistics were observed at intermediate distance classes (Figure 55), suggesting that fungal community dissimilarity is greatest between 650 km and 850 km. These correlations are comparatively weak and a dip in the locally weighted polynomial regression fitted to the left figure highlights this phenomenon; a reduced number of samples within this geographic range may be the cause.

As with the bacterial study, positive Mantel statistics were returned for the lower classes of environmental dissimilarity, after which negative values were observed. Mantel statistics were more negative for intermediate classes of environmental dissimilarity than those at greater dissimilarity, suggesting that fungal community dissimilarity was greatest at these intermediate values (Figure 56). A Mantel correlogram of environmental dissimilarity and geographic distance returned positive Mantel statistics for small distance classes up to approximately 40 km and negative Mantel statistics thereafter. Intermediate distance classes exhibited the most negative Mantel statistics, suggesting that environmental dissimilarity was greatest for these distance classes (Figure 57).

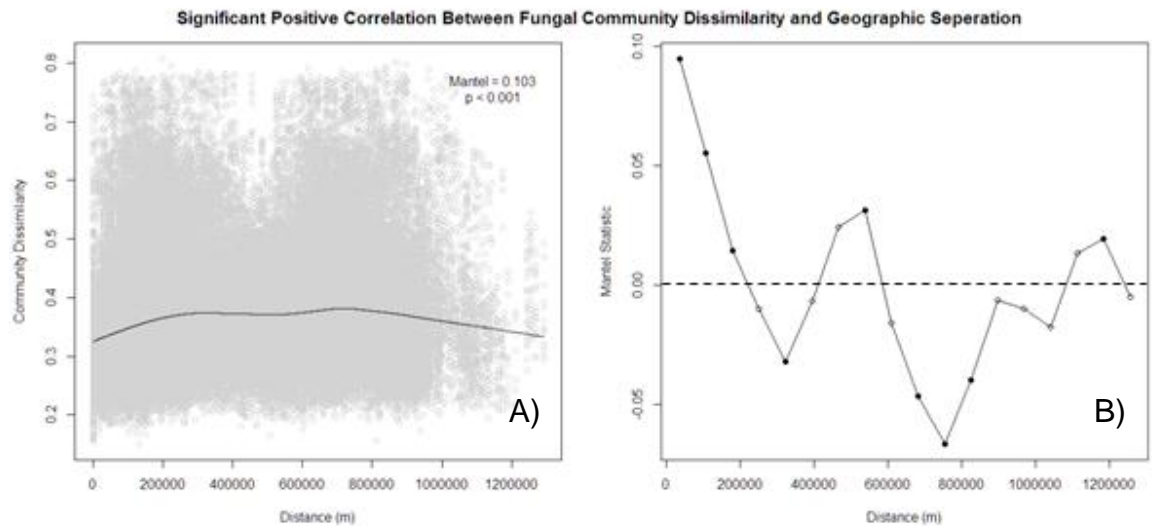


Figure 55: A) Mantel test showing that fungal community dissimilarity increases as geographic distance between samples increases (Mantel statistic = 0.103, $p < 0.001$). B) Mantel correlogram showing significant positive Mantel statistics are seen within distance classes representing geographical separation of sample sites below 200 km, between 450 km and 500 km and at 1150 km. Filled circles indicate values that are significantly different from zero; open circles indicate values that are not significantly different from zero.

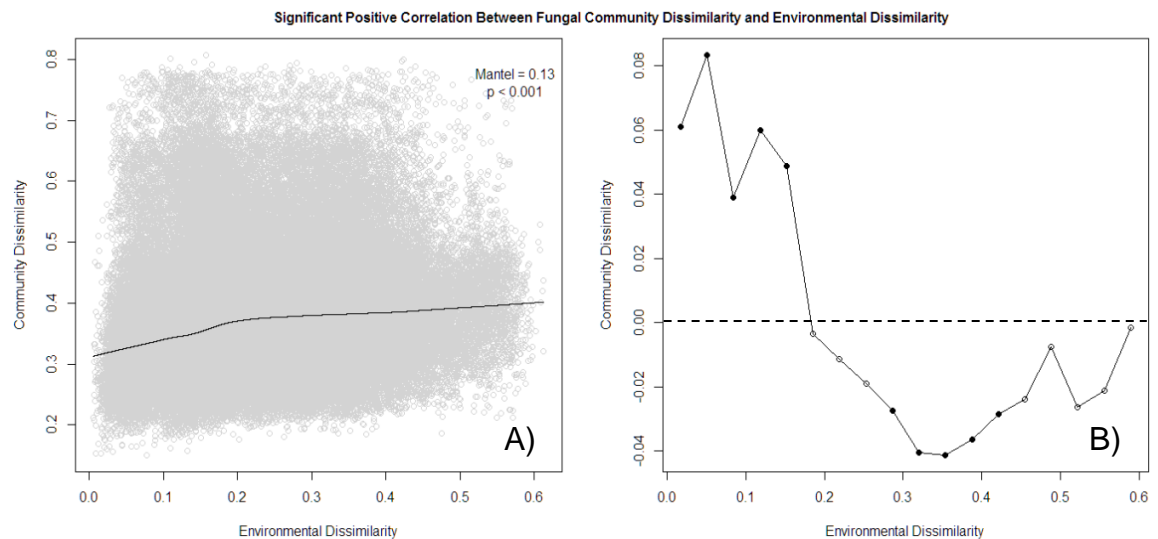


Figure 56: A) Mantel test showing that fungal community dissimilarity increases with environmental dissimilarity (Mantel statistic = 0.13, $p < 0.001$). B) Mantel correlogram showing that over short environmental gradients a positive Mantel statistic is seen in relation to fungal community structure, whereas from there on negative Mantel statistics are observed. Filled circles indicate values that are significantly different from zero; open circles indicate values that are not significantly different from zero.

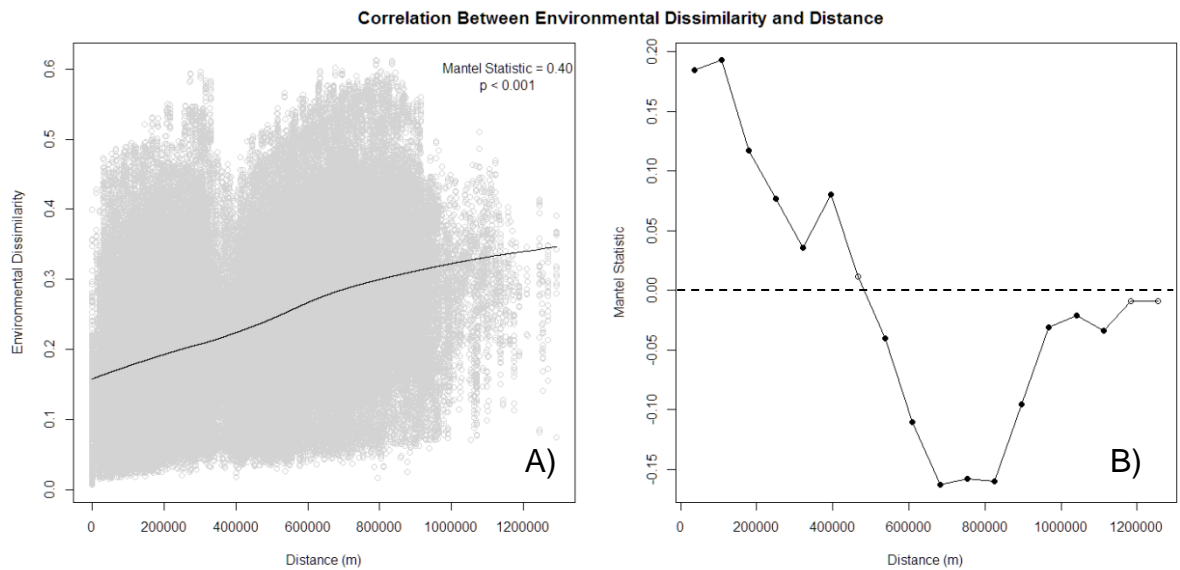


Figure 57: A) Mantel test showing that environmental dissimilarity increases with distance (Mantel statistic = 0.40, $p < 0.001$). B) Mantel correlogram showing the relationship between geographic distance and environmental dissimilarity. Filled circles indicate values that are significantly different from zero; open circles indicate values that are not significantly different from zero.

A partial Mantel correlogram was then constructed to show the correlations between fungal community dissimilarity and spatial separation of sample sites while controlling for the effect of the environment (Figure 58). At distance classes below 200 km significant positive spatial autocorrelation was observed.

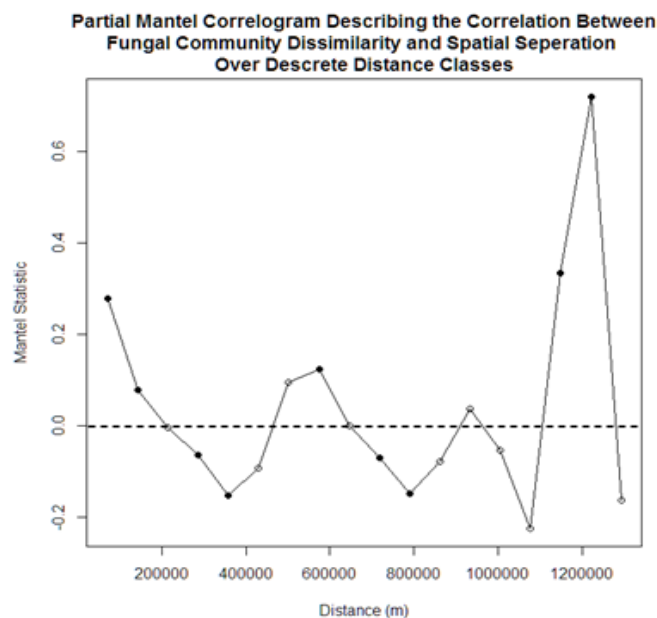


Figure 58: Partial Mantel correlogram showing spatial autocorrelation within soil fungal communities from samples within small and very large geographic distances of one another when the effect of environmental variability is removed. Filled circles indicate values that are significantly different from zero; open circles indicate values that are not significantly different from zero.

At intermediate distance classes, partial Mantel statistics fluctuate around the 0 point. Of most interest are the furthest distance classes representing samples between 1000 and 1200 km apart. Here very large significant positive Mantel statistics are observed. This was the largest Mantel statistic observed within any of the partial Mantel correlograms, suggesting that fungal communities separated by large geographic distances are more susceptible to spatial effects than any of the bacterial communities studied here. However, as with the bacterial spatial study, small sample numbers from locations separated by very large distances could potentially over inflate the Mantel statistic and thus results must be interpreted with caution.

7.4 Discussion

Here, a distance decay effect was observed in both bacterial and fungal populations, but in the case of the former, the magnitude and direction of this effect varies with the type of soil ecosystem under study. As some spatial autocorrelation occurs within these two kingdoms of life, parallels between the mechanisms which structure these populations and those of larger organisms may be drawn. This suggests that similar mechanisms limit or promote the distribution of all domains of life and affect community structure in similar ways. Due to this, other, more complex ecological theories derived from observations of larger organisms may now be investigated within microbial populations as many theories are reliant upon some form of spatial organisation to function.

7.4.1 Spatial Autocorrelation at Small and Large Geographic Distances

Several other studies have also noted a spatial effect upon bacterial community dissimilarity over short geographic distances, but this relationship falters at larger levels of spatial separation (King *et al.* 2010; Nunan *et al.* 2003; Nunan *et al.* 2002). Here, using a landscape scale approach, samples located at sites separated by very large distances (1200 km) exhibit a significant strong positive spatial autocorrelation, as well as those at smaller spatial scales. This Mantel statistic value of 0.4 is the highest observed within the bacterial partial Mantel correlograms, implying that positive correlations between community structures are strongest at the furthest distance classes but community dissimilarity is greatest at intermediate distance classes. However, the large Mantel statistic

reported for this distance class may be an artefact introduced by the small number of samples separated by large geographic distances. Dequiedt *et al.* (2011) suggest that local soil environmental characteristics are more important than spatial separation in structuring soil bacterial communities; they found that partial Mantel tests examining the correlation between community dissimilarity and environmental dissimilarity, whilst controlling for geographic space, resulted in large significant Mantel R values. Although this study (and findings in previous chapters) supports this hypothesis, spatial autocorrelation is still observed when environmental variability is controlled for, thus suggesting that although local scale environmental variability is the predominant driver of bacterial community structure, spatial separation is also an important factor. Furthermore, studies examining spatial autocorrelation in individual lineages of bacterial taxa, for example the pseudomonads, have shown a high degree of spatial organisation and suggest that global mixing of these communities does not occur (Cho & Tiedje 2000). If this is the case then, in contrast to claims by Dequiedt *et al.* (2011), soil bacterial populations do follow similar global scale patterns in biodiversity to larger terrestrial organisms.

Landscape scale assessments of soil bacterial spatial autocorrelation are comparatively rare. However, a study examining the degree of spatial autocorrelation in mountain habitats suggests that soil bacteria are highly spatially structured over distances up to 240 m (King *et al.* 2010). Here, high levels of spatial autocorrelation are also seen over relatively short geographic distances, suggesting that this level of spatial autocorrelation is applicable to multiple soil biomes. Further work examining the degree of spatial autocorrelation in tropical and arctic conditions may help devise a true global model of spatial autocorrelation within soil bacteria.

7.4.2 Community Variability and Taxonomic Resolution Affect Bacterial Spatial Autocorrelation

The assessment of spatial autocorrelation in the two subsets of soil samples representing distinct environments and bacterial communities suggests that the community variability and taxonomic level of assessment affect the degree of spatial and environmental correlations observed. Highly variable bacterial communities (e.g. those within low pH organic soils) exhibit stronger patterns of

spatial organisation than the more homogeneous populations found in the higher pH, *Lolium* dominated soils, which show no evidence for significant spatial autocorrelation. It can be theorised that, as all the *Lolium* dominated samples harbour similar bacterial communities, there is little variability across small or large geographic scales. However, the low pH *Calluna* dominated samples are highly variable, and a proportion of this variability can be explained by spatial separation. Therefore community variability plays an important part in determining the degree of spatial autocorrelation. Furthermore the degree of spatial autocorrelation found within the variable low pH communities is greater when examined at greater taxonomic resolutions, i.e. with DGGE rather than t-RFLP. The low taxonomically resolving t-RFLP method highlights weak patterns of spatial autocorrelation in the partial Mantel tests, but at levels approaching that of the species the degree of spatial autocorrelation increased. This suggests that specific habitats exhibit different patterns of spatial autocorrelation, and that species level spatial organisation may be greater than, for example, phylum level spatial organisation. Therefore the taxonomic resolution of the method chosen to examine spatial autocorrelation within microbial populations should be taken into consideration when an absence or presence of a spatial effect is reported.

Conversely, the species-environment relationships were stronger in both subsets when measured using t-RFLP, a method with low taxonomic resolution, than DGGE, suggesting that patterns relating to environmental variability at the landscape scale are somewhat different when examined closer to the species level than at less taxonomically resolved levels. The homogeneous bacterial populations residing within the higher pH *Lolium* dominated subsamples exhibited stronger links with the environment than the more variable populations residing within the low pH *Calluna* dominated samples, which were structured more in relation to geographic separation than environmental variability (defined by comparing Mantel R values).

7.4.3 Spatial Autocorrelation in Fungal Populations

Studies of species-area relationships within the fungal kingdom have largely been restricted to single types of fungal communities. Studies of ectomycorrhizal communities examined across habitats of different sizes within

forest habitats suggest that fungal species richness decreases as habitat size decreases and species-area relationships are similar to those generated for larger terrestrial organisms (Peay *et al.* 2007). Further evidence to support a species-area relationship within ectomycorrhizal communities is presented in Lilleskov *et al.* (2004), where small scale spatial patterns were also observed up to a maximum distance of 25 m. In contrast, when arbuscular mycorrhizal communities were examined over small spatial scales in wetland environments, no significant horizontal spatial autocorrelation was observed (Wolfe *et al.* 2007). Although the Ascomycota are globally distributed, evidence suggests that geographic distance is a more useful predictor of community turnover than the land system type from which they originate, and that there are predictable spatial patterns from the metre scale to the 100 km scale (Green and Bohanan, 2006). These studies suggest that spatial autocorrelation acts differently upon different fungal lineages and over different geographic scales. Within this study, examining total fungal community dissimilarity at the landscape scale, significant spatial autocorrelation was observed both at smaller scales (up to 15 km) and at very large spatial scales (>115 km), although small sample numbers from locations separated by large geographic distances are potentially problematic. Positive spatial autocorrelation at the larger distance classes was much greater than in the bacterial study, suggesting that dissimilarity of fungal populations is affected more by spatial separation than that of bacterial populations at the landscape scale. However, what must be taken into consideration is the relative taxonomic resolution of the two genes studied. As highlighted in Chapter 6, the pairwise sequence dissimilarity of the ITS region is greater than that of the 16S rRNA gene and thus, due to the faster rate of evolution of the ITS1 region (Nilsson *et al.* 2008), these two sequences do not analyse spatial patterns at the same taxonomic resolution. If this knowledge is combined with the findings of the DGGE based study, then the comparative taxonomic resolution of the community profiling method may inflate or deflate the apparent spatial effect observed and should be taken into consideration when reporting findings.

7.4.4 Comparisons with Macro Organisms

The mechanisms by which spatial effects occur within soil microbial communities is of great interest as they may be due to mechanisms that are

similar to or entirely different from those that cause spatial autocorrelation in macro organisms. For example, a principal and simple causative agent of spatial autocorrelation in macro organisms is a limitation to dispersal (Legendre, 1993). Individual bacterial and fungal propagules are small so parallels may be drawn between them and wind dispersed pollen of plants. Studies have shown that spatial autocorrelation in specific plant species is related to wind dispersal mechanisms (Dirnbock & Dullinger, 2004) and thus any bacterial and fungal populations utilising wind dispersal as a mechanism for distribution should, in theory, also be constrained by similar limitations. Others have argued that no limitation to dispersal is present within bacterial populations, principally due to their small individual size (Green & Bohannan, 2006). Ideally further work relating the comparative size and dispersal range of a microorganism may provide further insights into mechanisms of dispersal within soil bacterial populations. However, technical and ethical limitations to such experiments hinder progress. To elaborate on this, a way of tracking an individual microorganism through the environment would be to insert a non-natural marker gene via plasmid transformation prior to environmental release, as proposed in Kluepfel *et al.* (1993). However, releasing genetically modified organisms into the environment is controversial, and strict risk assessments are required (Hagedorn *et al.* 1994).

Specific fungal taxa have dispersal mechanisms linked to those of certain land mammals, for example through the ingestion of sporocarps and subsequent deposition of proliferating bodies (Janos *et al.* 1995). As many of the processes by which spatial autocorrelation occurs are observed in these land mammals (competition, social aggregation, predation, etc.), the effects of these processes may also indirectly affect the spatial distribution of the fungi. As an example, land mammals such as rats are known to aid in the dispersal of specific Basidiomycota taxa via the ingestion of sporocarps (Janos *et al.* 1995). Rats have been shown to exhibit many forms of social aggregation, be they via maternal behaviour or group aggression to intruders (de Bono, 2003) and thus the deposition of viable fungal bodies within faeces will be greater in areas where these carriers are located in higher abundances. This, theoretically, would result in a patchy distribution of fungal lineages belonging to this group, and an apparent spatial effect independent of environmental variability. Further

work detailing the correlations between spatial effects of vectors of fungal dispersal and fungal distribution itself within specific groups may aid in the understanding of microorganism distribution methods and potentially highlight unknown mechanisms by which spatial patterns of these organisms occur. Furthermore, many fungal taxa form macrostructures which proliferate through the soil substrate, also forming patchy areas of colonisation (Ritz & Young, 2004). Patchy areas of species distribution are a core concept of spatial ecology (Ettema & Wardle, 2002), but in this case, the patchy nature of hyphal forming fungi may be due to environmental constraints. Therefore studying such taxa individually over areas of similar environmental conditions may show distinct species area relationships in more detail than presented here.

7.4.5 What Can Cause Spatial Autocorrelation in Microbial Populations?

Other than limitations to dispersal, more complex interactions, primarily noted in macro organisms, may also be acting upon microbial populations. For example, social aggregation is seen in a multitude of terrestrial, freshwater and marine macro organisms, and similar processes could be said to occur in microbial populations through biofilm formation (O'Toole *et al.* 2000) or via the usage of co-metabolites (Horvath, 1972). Quorum sensing and cell to cell communication are present within a host of microbial populations and may be a driver of local spatial aggregation and distribution (Nunan *et al.* 2003). Therefore, small-scale but highly resolved studies into the spatial distribution of microorganisms within biofilms, and also the spatial distribution of cell to cell signalling molecules, may provide further insight into spatial autocorrelation at the local scale. A pioneering study into the spatial distribution of bacterial communities within the micro-architecture of soil has suggested significant spatial patterns under the millimetre scale and shown an increase in bacterial biomass close to pore spaces. Furthermore evidence for bacterial aggregation in cropped topsoils was also observed (Nunan *et al.* 2003). Spatial autocorrelation may occur from the very small scale to the large landscape scale, and these patterns are not independent (Franklin & Mills, 2003). Within this thesis, the km scale was predominantly examined due to the sampling design of the study. Future work examining spatial scaling from the very small (i.e. μm to mm) through to the metre and to the multiple km scale will allow a fuller understanding of spatial

autocorrelation within soil microbial populations at multiple levels of spatial resolution.

Microbial predation is another mechanism which may cause spatial patterns to emerge which are largely independent of the environment. Predation occurs in both bacterial and fungal populations via nematode interaction, phage host relationships, predation and parasitism of fungi by other fungal organisms or, bizarrely, bacterial mycophagy (Leveau & Preston, 2007). Predator-prey relationships are an interesting phenomenon within the soil microbial biome and are much understudied as a descriptor of microbial community structure at the landscape scale. As these processes have been shown to greatly affect the local and large scale distribution of macro organisms (Legendre, 1993), similar processes occurring within microbial populations may also result in spatial autocorrelation. Examples include nematode predation of bacteria and fungi, which is thought to greatly mediate fungal and bacterial biomass ratios (Wardle & Yeats, 1993), or conversely, fungal predation of nematodes (Mankau, 1980).

7.4.6 Does Functional Spatial Autocorrelation Exist?

Another interesting potential avenue for research is the spatial distribution and spatial autocorrelation of functional genes rather than taxonomic markers as reported here. Principally, if samples located closer together have a more similar functional gene portfolio than samples located at greater distances, the understanding of the distribution of functional genes within the environment, and the mechanisms by which the distribution is mediated, may be improved, which would help further the concept of ecosystem services at the gene level.

7.4.7 Is It Really Spatial Autocorrelation?

An interesting topic of discussion is the independence of spatially mediated descriptors of microbial communities from environmental variability. All spatial statistical methods used to elucidate spatial autocorrelation within any population assume that a “pure spatial” effect can be found by partitioning the variability due to the environment and to space itself in the variability in community structure of a set of organisms. However, any mechanism by which spatial autocorrelation may occur seems to be in some way linked, no matter how tenuously, to the environment. For example a study by Dirnbork *et al.*

(2004) found significant spatial autocorrelation in all taxa of alpine plants studied, and suggests that spatially structured processes occur independently of the environment. The authors go on to propose hypotheses to describe these spatially structured processes, and suggest that spatial autocorrelation within alpine plant communities forms as a result of the differences in dispersal ability. Furthermore, plants with small diaspores, adapted for wind dispersal, show higher levels of spatial autocorrelation than those adapted for other methods of dispersal. However, evolution of small diaspores for wind dispersal is partly due to environmental variability, and the mechanism by which dispersal occurs is directly related to the environment itself (i.e. wind strength) (Jongejans & Schippers, 1999). It may be argued that it is difficult to find a mechanism by which spatial patterns are generated which is not linked in some way to environmental variability, and therefore apparent spatial autocorrelation occurs due to unmeasured environmental variability. This point is addressed in a landmark paper by Borcard *et al.* (1992) addressing the effects of spatial separation upon oribatid mites within an undisturbed sphagnum moss ecosystem. This paper suggests that, although the unexplained proportion of community variability is a separate entity, it is likely to comprise unmeasured spatial and environmental characteristics. Furthermore, the unexplained variability shows the need for more complex functions to explain the variability in the community matrix. As an example, they describe unmeasured micro-scale variability in nutrients such as ligneous matter, or the distribution of fungi, algae, and pollen, as potential determinants of spatial structure at small scales. Therefore measurement of spatial and unexplained variability is a useful tool to try to discern mechanisms by which communities are structured (as in Dirnbork *et al.* 2004).

7.4.8 Conclusions

In conclusion spatial separation has a distinct effect upon both fungal and bacterial populations and, to a degree both sets of organisms conform to Tobler's first law of geography (Miller, 2004). Due to this it can be said that these organisms are structured in similar ways to macro organisms. However, the mechanisms by which spatial autocorrelation occurs are unknown. Furthermore, as a spatial effect is present within soil microbial populations at the landscape scale, the next aim is to determine the mechanisms by which

these spatial patterns are generated. Within communities of macro organisms processes such as competition, social aggregation, and competition partially drive these patterns, and thus these processes may also be important within microbial communities and should be investigated.

The taxonomic resolution used to assess microbial community structure affects the degree of spatial organisation reported, and thus it is difficult to truly compare the degree of spatial organisation between different microbial kingdoms. 16S rRNA gene targeted t-RFLP approaches describe spatial effects at a comparatively low taxonomic resolution (i.e. approximately phylum) whereas DGGE and ITS targeted t-RFLP describe spatial patterns arising at closer to the species level. Therefore it is important to ascertain at which level of taxonomic resolution spatial patterns are being described.

7.4.9 Future Work

Further analyses upon these data sets may be carried out using more complex methods of discerning spatial autocorrelation. A proposed method known as principal coordinates of neighbourhood matrices (Dray *et al.* 2006) may provide a greater insight into the variation due to spatial separation as well as highlighting variation in the community dissimilarity matrices of unknown origin. Also, datasets generated here may provide ideal examples for the construction of further spatial models specifically targeted towards microbial populations by other research groups more versed in the construction and use of complex spatial statistics.

An interesting topic of discussion is whether ITS based assessments of spatial autocorrelation within soil fungal communities are directly comparable to assessments of spatial autocorrelation in other organisms using different taxonomic markers. As highlighted in Chapter 6, the pairwise dissimilarity of ITS sequences is greater than that of 16S rRNA sequences and thus, due to the greater level of sequence variability, it is hypothesised that these two sequences are not examining ecological patterns of soil microbial communities at the same taxonomic resolution. Therefore, although it has been suggested here that soil fungal communities exhibit a greater degree of spatial autocorrelation than soil bacterial communities, these spatial patterns may relate to completely different taxonomic levels. Comparisons of fungal spatial

autocorrelation using 18S rRNA genes and ITS genes may identify differences in the degree of spatial autocorrelation observed within this kingdom.

Chapter 8: Final Discussion

“Anyone who has studied the history of science knows that almost every great step therein has been made by the “anticipation of nature”, that is, by the invention of hypotheses, which, though verifiable, often had very little foundation to start with; and, not un-frequently, in spite of a long career of usefulness, turned out to be wholly erroneous in the long run.”

Thomas Henry Huxley 1887

In ‘The progress of Science 1837 - 1887’ *Collected Essays* (1901)

8.1 Overview of Thesis

The principal aim of this thesis was to gain insights into the ecological factors which affect the diversity, community structure, and community composition of two distinct lineages of the soil microbiota, namely soil bacteria and fungi. Soil bacteria and fungi were chosen as model organisms to study in relation to soil microbial biogeography as these domains of life exhibit very different physical (Madigan *et al.* 2003), evolutionary (Hurst *et al.* 2002), and chemical (Hammel, 1997) life processes, while occupying similar physical niches (at least at the landscape scale) (Boer *et al.* 2005). Bacteria are unicellular, prokaryotic, and comparably ancient within the soil environment (Doolittle *et al.* 1996), whereas fungi are eukaryotic, physically different in construction (Lipke *et al.* 1998) and in many cases multicellular with different life cycle stages (Hurst *et al.* 2002). Soil bacteria and fungi are the most dominant members of the soil microbiota (Bardgett, 2005), and fulfill a wide range of important ecosystem functions (Beare *et al.* 1995). Identifying landscape scale patterns in soil microbial biodiversity helps to increase understanding of how these communities respond to environmental gradients, provides information relevant to soil microbial ecological theory, and gives a stable platform on which to base further research into microbially mediated, soil based, ecosystem services.

The UK has a temperate landscape with diverse geology (Bennison & Wright, 1969), varied climate (Arnell *et al.* 1990), and a huge array of soil types over a comparatively small geographic area (Haygarth *et al.* 2009), thus making it a perfect model to analyse soil microbial biogeography. In order to assess

landscape scale patterns in soil microbial biodiversity, a large number of soil samples were required, representing the huge diversity of soils and habitats present within the UK. The Countryside Survey 2007 (Carey *et al.* 2008) was ideal in this regard as it provided over 1000 soil cores from a statistically relevant sampling design with detailed accompanying physicochemical, environmental, and spatial data for each soil core. It was these soil samples which formed the basis of the work detailing patterns in soil microbial biogeography and thus, this thesis could not have been attempted without the input of a large number of people collecting the diversity of soils found within the UK.

8.1.1 Optimising a Soil Nucleic Acid Extraction Method

The next step in generating a landscape scale picture of the microbial biodiversity of the UK was to optimise a method for the extraction of genetic material from a wide variety of soil types to provide a basis for the molecular analysis of the microbial communities contained within them. In order to do this, an existing nucleic acid extraction procedure (Griffiths *et al.* 2000) was optimised and assessed not only for use with multiple soil types, but also for potential biases introduced into bacterial community profile analysis by incomplete extraction of genetic material. It is well documented that soil physicochemical properties can not only affect the recovery of nucleic acids (Miller *et al.* 1999) due to physical entrapment (Taylor *et al.* 2002) or charged interactions (Cai *et al.* 2006), but can also affect the purity of nucleic acid extracts via the co-extraction of contaminants such as PCR-inhibiting humic acids (Jizhong *et al.* 1996; Smalla *et al.* 2008). Feinstein *et al.* (2009) showed that incomplete extraction of nucleic acids from soil samples can potentially affect t-RFLP based assessments of bacterial community structure (but not fungal) and thus this was also assessed. Modifying the original method by using a reduced sample mass of 0.25 g (from 0.5 g) and adding a CTAB based freeze thaw step gave an increased nucleic acid yield from clay soil types (previously producing low nucleic acid yields) and allowed PCR amplification of target DNA from a wide variety of soil types. Furthermore when the effect of incomplete extraction was assessed across multiple soil types, t-RFLP based bacterial community profiles generated from a single extraction did not exhibit contradictory patterns to those generated from multiple pooled extracts. This

showed that the method was suitable for high throughput, landscape scale assessments of microbial biogeography encompassing a wide range of soil types. Although the reduction of sample mass from 0.5 g to 0.25 g was assessed here, the use of much larger sample masses (10 g – 100 g) was not and thus is an avenue for further research. Using larger quantities of soil not only reduces the biases associated with the general heterogeneous nature of soil across small areas, but may reduce biases towards particular taxa or kingdoms. This may be especially important when fungal targeted studies are undertaken as fungal macrostructures, such as hyphae, may not be included if smaller quantities of soil are used. Furthermore, plant and root structures are less likely to be included if small sample sizes are used, potentially resulting in an under representation of root associated organisms in the final analyses. Bead mill lysis is regarded as the best cell lysis mechanism available (Miller *et al.* 1999), and thus physical limitations occur when trying to extract nucleic acids from large quantities of soil due to the size of available apparatus. As 0.25 g was found to be an optimal mass of soil to extract nucleic acids from, multiple technical replicates would be needed in order to increase the mass of soil from which nucleic acids were obtained. In the case of small scale studies, this may be of interest if the shifts in community composition observed are affected by the quantity of soil analysed per sample. However when performing large scale studies, using thousands of samples, extracting nucleic acids from large quantities of soil per sample greatly increases time and financial expense.

8.1.2 Broad Scale Patterns of Soil Bacterial Community Structure are Highly Related to Environmental Variability

From the national archive of soil nucleic acids described in Chapter 2, patterns of soil microbial biodiversity were analysed across the UK using an array of molecular based methods. First a large scale t-RFLP based assessment of soil bacterial communities was carried out and showed that numerous environmental characteristics determined the community structure and composition of these organisms. More specifically, soil pH was shown to be a major determinant of soil bacterial diversity and overall community composition, although the above ground plant communities and soil nutrient status were also strongly related as well as other environmental variables to a lesser degree. Low pH organic soils were dominated by a few bacterial OTUs, which were

identified as belonging to Acidobacteria lineages through *in-silico* t-RFLP analyses of a global database of 16S rRNA sequences. As soil pH increased, alpha diversity also increased, with higher pH improved habitats being categorised by proteobacterial lineages (predominantly Alphaproteobacteria) and actinobacterial taxa. Beta diversity was assessed and, although low pH organic environments showed a reduced alpha diversity, beta diversity was much greater than that of higher pH soils, indicating a larger degree of variability of bacterial community structure among sites with low pH soils. This may be because the low pH soils exhibit a greater degree of variability than medium and high pH soils (Chapter 3). However, when alpha and beta diversity scores were added together, an overall upward trend in gamma diversity with pH was found. Interpolated mapping of soil bacterial community structure revealed patterns relating to soil gradients across the country, and areas with similar soil and climatic properties produced similar bacterial t-RFLP profiles. Multivariate partitioning of bacterial community profiles based upon reducing the community dissimilarity of partitioned profiles in relation to shared environmental characteristics indicated that soil bacterial communities are structured in relation to environmental variability, and a basic predictive model was constructed describing this relationship. Although t-RFLP allows the identification of overarching patterns in microbial biodiversity across a landscape (Fierer, 2008), it is not taxonomically informative and has an uncertain level of taxonomic resolution. Therefore pyrosequencing was undertaken to assess compositional differences among soils with different physicochemical properties highlighted as important in determining bacterial community structure and diversity. Furthermore, these patterns were assessed using two separate regions of the 16S rRNA gene in order to define the effects of gene region choice upon findings.

8.1.3 Pyrosequencing Sheds more Light upon Changes in Bacterial Community Composition with Relation to Environmental Variability

Pyrosequencing analysis revealed that the soil conditions highlighted as important bacterial community determinants in Chapter 3 had consistent effects when sequencing approaches were used, irrespective of the 16S rRNA gene region targeted. Pairwise dissimilarity matrices constructed from V1 – V3 region and V6 – V9 region sequences showed that the V1 – V3 region of the 16S

rRNA gene was more variable than the V6 – V9 region, supporting the hypothesis that the V1 – V3 region can better discriminate between closely related bacterial taxa (Neefs *et al.* 1990), due to the fast rate of evolution inherent to the V1 region (Schloss, 2010). Furthermore this explains why a larger number of OTUs were generated at 97 % similarity when the V1 – V3 region was analysed than with the V6 – V9 region. Diversity (Chao1 richness) was lower within the low pH soils than the medium and high pH soils, corroborating findings presented in Chapter 3, and these patterns were consistent irrespective of the region of the 16S rRNA gene targeted. However, compositional differences at the phylum level were observed between the two regions sequenced. Proteobacterial, acidobacterial and actinobacterial sequences predominantly comprised those generated from sequencing the V1 – V3 region of the 16S rRNA gene, whereas over 60 % of the sequences returned from the V6 – V9 region analysis were identified as acidobacterial taxa. As these three phyla are the most dominant in soils (Janssen *et al.* 2006) these results highlight the need for a standardised approach to pyrosequencing studies of soil bacteria so that comparisons can be accurately made between studies.

Although phylum level differences were observed, the general changes in community composition across the environmental gradients studied were comparable when either gene region was used. Acidobacteria group 1 dominated the organic low pH soils while Acidobacteria group 6 was present at high abundances in high pH improved soil habitats. It was hypothesised that the large pH effect observed in Chapter 3 was predominantly due to the changes in the proportional abundance of Acidobacteria group 1 and Acidobacteria group 6 across environmental gradients. Furthermore, when indicator taxa analysis was performed upon sequences generated from either region of the 16S rRNA gene studied, acidobacterial subgroups were the predominant lineages identified as significant indicators of soil pH. Alphaproteobacteria did not follow the unimodal distribution with pH proposed in Chapter 3; however, Betaproteobacteria were shown to be indicator taxa of the medium pH soils under study. Phylogenetic trees revealed that OTUs identified using the V1 – V3 region sequences clustered together with regard to assigned taxonomy, but OTUs generated from the V6 – V9 region sequences were dispersed across the tree. This suggests

that the V1 – V3 region is preferable to the V6 – V9 region for accurate taxonomic assignments.

Further analysis of OTU sharing between and within the pH groups using the phylogenetic trees shows that low pH samples contained a high number of unique OTUs while medium and high pH samples shared more OTUs, although OTU sharing across all samples was minimal. This corroborates findings presented in Chapter 3 suggesting that the community composition of low pH soils is more varied than that of medium and high pH soils (i.e. beta diversity). When patterns in bacterial biodiversity were analysed closer to the species level, areas with similar physicochemical properties displayed different patterns of biodiversity as highlighted by the detection of different OTUs within replicates of the same soil groupings. Therefore studies with higher taxonomic resolution suggest that soil bacteria conform to biogeographic principles derived from observations of macro organisms.

Distinct patterns in bacterial biodiversity were seen at the landscape scale using t-RFLP and these patterns were shown to be highly related to environmental gradients, specifically soil pH. Compositional changes were identified across these environmental gradients using pyrosequencing, showing that biogeography is relevant to soil bacterial populations. The next task was to determine if other dominant members of the soil microbiota, namely soil fungi, exhibit similar patterns of biodiversity at the landscape scale, and if not, to determine the factors which control the distribution and community structure of this microbial kingdom.

8.1.4 Broad Scale Patterns of Fungal Biodiversity are not Similar to Those Observed in Bacterial Populations, but may be Describing Patterns at Different Taxonomic Resolutions.

Large scale ITS based t-RFLP showed that fungal diversity was high within every sample and evenness was low, supporting findings in other studies suggesting that the variability in fungal communities between soils is high (Fierer *et al.* 2007; Buée *et al.* 2009). Fungal communities did not respond to environmental gradients in the same manner as bacterial populations when t-RFLP profiles were analysed, but instead categorical variables such as the dominant plant species were shown to be the most important determinants of

fungus community structure. However, beta diversity assessments of fungus communities using the t-RFLP method showed that a reduction in beta diversity occurred as pH increased, indicating that trends in community variability were similar in both bacterial and fungus populations in relation to increasing soil pH. Methodological limitations to the ITS t-RFLP study had to be taken into consideration when analysing results, including the fact that inter and intra species variability of this sequence is high (Avis *et al.* 2005; Kodie *et al.* 2005), and that the taxonomic resolution of the ITS t-RFLP approach may be analysing patterns of biodiversity at different levels to that of the bacterial 16S rRNA gene t-RFLP approach (Chapter 5). This was highlighted when small scale clone library analysis was undertaken and distinct patterns in community composition and diversity were seen across environmental gradients. When OTUs were clustered at lower levels of sequence similarity, low pH organic soils exhibited a lower Simpson's diversity than medium and high pH improved soils, consistent with the diversity trends noted for bacterial taxa. PCA plots representing fungus community dissimilarity showed that the community variability of the low pH organic soils was greater than that of the medium and high pH soils, implying that beta diversity is greater in low pH organic environments than medium and high pH improved soils, again similarly to patterns in diversity seen with bacterial communities. OTU overlap was minimal when OTUs were clustered at 97 % similarity, concurring with findings presented in Fierer *et al.* (2007) that fungus OTU evenness is low. When OTUs were clustered at lower levels of sequence similarity, medium and high pH samples shared more OTUs with each other than either did with the low pH samples, suggesting that soils of more similar physicochemical composition are more similar in fungus community composition and thus environmental variability plays an important role in structuring soil fungus communities. Compositional differences were seen between soil groups of contrasting physicochemical properties, the most obvious of which was the dominance of Basidiomycota lineages, predominantly Agaricomycota, in low pH organic environments. Agaricomycota lineages were in low abundance within the improved soils of the medium and high pH samples, where members of the Basidiomycota class Tremellomycetes were seen in higher abundances. Although the proportion of sequences defined as Ascomycota was reasonably consistent across all samples, class level patterns were observed in relation to the environmental gradients analysed. A large

proportion of the Ascomycota sequences within the low pH organic soils was shown to belong to the class Leotiomycetes, which decreased in proportional abundance in the medium and high pH samples where it was replaced by a dominance of the Sordariomycetes. High pH improved soils were characterised by Chytridiomycota lineages, which were largely absent from the low pH organic soil types as were Zygomycota OTUs. A phylogenetic tree was constructed which showed that sequences originating from low pH soils clustered together in large clades, predominantly with reference to the sample of origin, whereas sequences from medium and high pH samples showed little within group or sample clustering. This showed that individual low pH organic soils had little sequence diversity but the variability among low pH samples was high, providing further evidence to suggest that beta diversity is higher in low pH organic soil systems than medium and high pH improved soil habitats. As ITS t-RFLP and clone library analyses exhibited conflicting results in terms of taxa-environment relationships, further analysis was undertaken on a larger number of soil samples using pyrosequencing. This study targeted a gene with lower intra species variability than the ITS gene, and more comprehensive, open access sequence repositories, namely the 18S rRNA gene.

8.1.5 Fungal Pyrosequencing Confirms an Environmental Effect upon Community Assembly

Using pyrosequencing, similar environmental variables highlighted as important in structuring bacterial communities were seen to be important as factors affecting soil fungal community structure. These were the above ground plant community and soil pH. However, the relative importance of these variables was reversed in the fungal based study, with the above ground plant communities exhibiting larger R^2 values than soil pH when examined via least squares regression fits upon NMDS ordinations. NMDS ordinations and PCoA plots of Unifrac dissimilarity matrices showed that samples separated out across the plots in reference to the pH group of origin in a similar fashion to that seen in the bacterial based pyrosequencing study. However, UPGMA trees showed that low pH organic samples were different in fungal community composition from the medium and high pH samples, but medium and high pH groups were comparatively similar. The dominant fungal phyla seen in each pH group were not consistent with the results of the small scale clone library

analysis using the ITS gene. Ascomycota dominated low and high pH groups, while Basidiomycota dominated medium pH groups. This implies that sampling depth and taxonomic marker gene of choice have significant effects upon the reported community composition and thus must be taken into consideration when comparing studies using different sequencing approaches. However, class level patterns were relatively consistent between the studies.

Leotiomycetes lineages made up a large proportion of the low pH Ascomycota taxa, while a shift towards a dominance of Sordariomycetes was seen in the medium and high pH groups. Agaricomycota were in high abundances within the low pH samples, which was consistent with findings in Chapter 5, but tremellomycete dominance was only observed in the medium pH samples and not in the high pH samples as seen in Chapter 5. With reference to findings in other studies, it was hypothesised that the Leotiomycetes were the most sensitive to changes in soil pH, while other taxa had more defined links with other soil physicochemical characteristics. Changes in the proportional abundance of this class across the environmental gradients under study were hypothesised to be predominantly responsible for the pH effect seen in the soil fungal communities, and parallels could be drawn between this lineage and the acidobacterial subgroups highlighted as those predominantly affected by changes in soil pH in the bacterial pyrosequencing study (Chapter 4).

Phylogenetic trees constructed from pyrosequencing of the 18S rRNA gene showed that OTU sharing between samples was small, even at lower levels of OTU clustering similarity. OTU sharing within pH groups was greater than that between groups, again showing that environmental variation is an important determinant of fungal community structure.

Although strong relationships between environmental conditions and the microbial communities under study were found, the effect of geographic isolation was examined to further explore the effect of geography upon landscape scale patterns of biodiversity within the bacterial and fungal kingdoms.

8.1.6 Soil Bacteria and Fungi Show Evidence of Spatial Organisation, but Taxonomic Resolution Affects the Magnitude of this Phenomenon

The large scale t-RFLP analysis of fungal and bacterial community structure was used as a basis to explore geographic effects upon these two kingdoms, and in either case, spatial autocorrelation was observed. This shows that although environmental characteristics are the principal determinants of bacterial and fungal communities, spatial effects are also important. Within the partial Mantel correlograms describing the correlation between geographic separation and community dissimilarity while controlling for environmental dissimilarity, significant positive spatial autocorrelation was observed at small and very large scales within both kingdoms. These values were larger for fungal communities than they were for bacterial communities, suggesting that at small and large distance classes the effect of spatial autocorrelation is greater within fungal communities than bacterial communities. However, the taxonomic level at which ITS and 16S rRNA gene targeted t-RFLP assess spatial organisation in soil microbial communities is not consistent and so these approaches may be describing spatial patterns at differing taxonomic levels. Furthermore, bacterial targeted DGGE analysis showed that the taxonomic resolution of the chosen community profiling method alters the degree of correlation between community structure and environmental variation, as well as the strength of observed spatial effects. Highly variable communities from low pH organic soils showed a greater degree of spatial organisation than the comparatively homogeneous populations found within the higher pH soils and thus the relative effect of spatial autocorrelation was related to the type of bacterial community residing in each soil habitat. As significant spatial autocorrelation was observed in both kingdoms, it can be hypothesised that, although “the environment selects” it is not the only factor which needs to be taken into consideration when determining the major influences upon soil microbial community structure and composition, and that other factors which affect the global distribution of macro organisms may also play strategic roles in the community assembly of these diverse and important microbial populations.

8.2 Impacts upon Soil Microbial Ecological Theory

Following early studies into general biogeography, a fundamental law was proposed by Candolle and Sprengel that “...the lower the organization of the

body is, the more generally it is distributed. As infusory animalculae are produced in all zones, when the same conditions exist; we find the same manner that Fungi, Sponges, Algae and Lichens ... are distributed upon the earth, in the sea, and in the waters, when the same circumstances propitious to their production occur" (translated in O'Malley, 2007). However, when bacterial and fungal communities were assessed by pyrosequencing, numerous OTUs generated at 97 % similarity were unique to particular pH groups and samples within these groups. Although even the most up to date sequencing methods do not capture the entire diversity within a sample (Kunin *et al.* 2010), these results provide evidence that not all bacterial species are everywhere and thus this postulated fundamental law may be intrinsically flawed. Furthermore, other studies have noted endemic microbial taxa such as pseudomonads in soils (Cho & Tiedje, 2000) or *Synechococcus* taxa in hot spring microbial mats (Papke *et al.* 2003). Although this law may not be wholly accurate, comparisons between bacterial and fungal spatial autocorrelation in Chapter 7 show that the effect of spatial auto correlation within fungal communities may be greater than within bacterial communities, implying that the ability to disperse may, in some form, be related to organism complexity. Although the above quotation was originally intended as a comparison between macro and microbial biogeography, a similar hypothesis can be applied to the microbial systems under study here. Many fungal lineages form macro structures, be they hyphal masses or fruiting bodies (Madigan *et al.* 2003). This means that fungi cross the boundary between the microbial and the macro-bial worlds. If this is truly the case spatial effects seen within fungal organisms with life stages of this nature may be similar to those seen in plants. However, to truly address the relations between organism complexity and biogeographic patterns, detailed comparative studies into the effects of spatial autocorrelation upon microorganisms of the same kingdom forming macro structures and those that are unicellular in nature are needed. Coupled with this, further analysis of spatial organisation described by methods with different taxonomic resolution is needed before direct comparisons of spatial relationships can be made between kingdoms.

Communities of soil bacteria predominantly seem to be structured more by environmental conditions than spatial separation, whereas the opposite is true for soil fungi (Chapters 3-7). Both kingdoms have been shown to be locally as

well as globally diverse (Fierer *et al.* 2007). However, bacteria are older than fungi in an evolutionary context (Hugenholtz & Pace, 1996). This suggests that the large diversities inherent to both kingdoms evolved over different evolutionary time scales. If speciation/extinction rates are the principal drivers of diversity by definition (Mittelbach *et al.* 2007), they may be somewhat different between these two kingdoms. The number of soil fungal OTUs has been shown, in some instances, to exceed that of bacterial OTUs within the same soil environments (Fierer *et al.* 2007). Mechanisms of speciation within these two kingdoms are fundamentally different, or act at different scales and rates. If fungal communities are structured more by the influence of spatial separation, because they form macrostructures, than bacterial communities, which are assembled more in relation to environmental variability, it may be the case that allopatric speciation occurs more frequently in populations of the former than the latter. This would, in principle, explain the larger numbers of fungal OTUs detected within the soil ecosystem, as well as the faster rate of evolutionary radiation. Of course, direct comparisons of diversity indices between the two kingdoms using single genes with different levels of variability cannot be made with conviction, but OTU based assessments give researchers a baseline for comparison in this regard.

Soil pH was highlighted throughout the chapters as a major factor influencing biogeographic patterns in bacterial communities, and to a lesser degree fungal communities. In the case of the former the effect of pH was most prevalent, being the principal determinant of bacterial community structure and diversity. The effect of soil pH upon fungal communities was only really observed using directed sequencing approaches in Chapters 5 and 6, where the taxonomic resolution of results could be altered. Although pH was highlighted as an important factor, it is still unclear whether pH is a direct or indirect factor, as pH may cause biogeographic patterns due to direct intracellular stress or via the modulation of unmeasured processes. Examples of these processes may include alterations in the bioavailability of certain compounds to organisms due to changes in soil pH (Dalal *et al.* 2001; Bardgett, 2005), or the fact that certain taxa may be sensitive to pH and thus community assembly may be altered because of the absence of organisms producing co-metabolites essential for the proliferation of other taxa. Furthermore, it is also unclear if soil pH acts in a

similar manner upon the two kingdoms. Bacteria may be more sensitive to fluctuations in pH due to differences in cellular construction: possession of “naked” DNA (Madigan *et al.* 2003), differences in cytoplasmic barriers to the external environment (Lipke & Ovalle, 1998), or susceptibility to toxic metal release (Pietri & Brookes, 2008). A study on the soil from the Rothamsted, Hoosfield long-term pH manipulation experiment suggests that low microbial biomass and diversity within acidic soil are related to an increased level of extractable aluminium (Rousk *et al.* 2009). Aluminium inhibits specific microbial processes (Pietri & Brookes, 2008) and this toxicity imposes a selection pressure on the population. Conversely, the proliferation of specific taxa at the extremes of pH may indirectly affect the diversity of the system by promoting or inhibiting the growth of other organisms via the secretion of co-metabolites or anti-microbial agents (Hibbing *et al.* 2009). All these hypotheses directly relate to other aspects of biogeography, namely competition, predation, and extinction. The fact that the environment and spatial separation do play specific roles in structuring these microbial populations implies that processes such as competition, predation, and extinction are occurring (Legendre, 1993). For an organism to proliferate within certain habitats at the expense of others suggests that it is more suited to exploiting a particular niche than those that are at low abundance or below the detection limit of the chosen method of study (Hardin, 1960). Within the soil microbial world this can be seen most clearly in the case of the group 1 Acidobacteria in low pH soil environments. T-RFLP and pyrosequencing analysis shows that low pH, organic soils are unique in their community composition, with large proportions of the entire community being members of this group. Moreover, members of this group appear to be reasonably restricted, in terms of proportional abundance, to these soil types, suggesting that they operate with a competitive advantage over other taxa within these systems. Although, as a phylum, true endemism does not occur, patterns implying an endemic nature within OTUs classified towards the species level begin to become apparent. As an entire phylum, the Acidobacteria do in fact seem to be widespread in distribution within the soils under study here. As widespread distributions of bacterial and fungal taxa are reliant upon the ability to be widely dispersed and to proliferate within a wide range of ecosystems (Ramette & Tiedje, 2007), it is interesting that members of lineages of similar

descent can be found within habitats of contrasting environmental conditions (in particular the group 1 and group 6 Acidobacteria).

Acidobacteria are generally slow growing and long lived, with slow metabolic rates (Ward *et al.* 2009), and thus fulfill the criteria of K strategists (Madigan *et al.* 2003). As different subgroups are found in areas with different pH, it is perhaps not just the slow growing nature of these organisms which gives them the selective advantage within low pH soils. Bacterial phyla which are dominant in medium and high pH soils (such as the Alphaproteobacteria) are generally fast growing with short generation times under conditions of excess nutrients (Smit *et al.* 2001), potentially allowing them to out-compete the slower growing Acidobacteria in soils of this nature. The R and K selection theory is generally regarded as an over simplification, but may provide a useful framework for further research in this area when the ecological characteristics of different microbial taxa are compared (Fierer *et al.* 2007).

According to clone library analysis and pyrosequencing of fungal communities across a pH gradient, a restriction in diversity is also seen, suggesting that organisms of specific taxonomic descent have also adapted to survive in conditions not optimal for cellular function. Although the division is not as clearly defined as the acidobacterial subgroup 1 and 6 divide, specific fungal taxa were shown to be statistically linked to specific areas with different soil pH (Chapter 6). Coupled with this, high rates of co-evolution between plants and specific soil fungi (Isaac, 1992) have resulted in biogeographic patterns relating, to a degree, to the above ground diversity. Again, specific fungal taxa were only detected within areas with specific environmental conditions, in particular the Chytridiomycetes and Leotiomycetes. Thus, both bacterial and fungal populations seem to adhere to Gause's law of competitive exclusion (Hardin, 1960), at least within OTUs close to the species level. Furthermore, areas with the same environmental conditions show different patterns of biodiversity within the two kingdoms, suggesting that, at taxonomic levels close to that of the species, Buffon's law of biogeography is in action (Cox & Moore, 2010). Therefore not only are overarching patterns of soil microbial biogeography similar to those of larger organisms, but more complex hypotheses on the generation of macro ecological patterns seem to fit within the context of the soil microbial world.

8.3 General Criticisms

Empirical falsification is the primary means by which most scientific studies aim to draw conclusions, that is by the generation of hypotheses which, through experimental assessment, can either be supported or refuted (Popper, 1959). Within the field of biogeography as a whole, this is fundamentally difficult as biogeographical studies principally aim to simplify the complex natural world in an effort to elucidate patterns which provide a foundation for further research. Hypothesis testing is reliant upon *a priori* knowledge of the system under study which, in cases such as these, is largely non-existent. Inductive reasoning, such as that used by Darwin in his assessments of macro-bial biogeography, break from this school of thought and have thus been labeled by some as unscientific (although this claim was later partially recanted) (Hull, 1999). If this is the case then, by definition, this body of work also cannot be viewed as classical scientific experimentation, and as such the word “survey” has been used throughout. To use null hypotheses such as “there are no biogeographic patterns within soil microbial populations” seems somewhat forced in an effort to adhere to the empirical falsification approach. Furthermore, although challenging the “Everything is everywhere” hypothesis may aid in experimental design, it has been vehemently argued that this theory cannot be called a hypothesis as it is inherently untestable (Purdy, K. 2012, personal communication). Here hypotheses and null hypotheses were only generated and tested in hindsight, after initial data mining and multivariate analyses had taken place. However, as with Darwin’s inductive reasoning on the origin of species (Darwin, 1859), approaches such as those presented here are essential in emerging disciplines as they provide a basis for true scientific experimentation to take place to either confirm or refute hypotheses. Simply, we have to start somewhere.

A potential question about the studies presented here is how an understanding of soil microbial biogeography can be applied. One principal advantage of this study is the fact that the soil samples which form the basis of microbial community profiling come from a continuous survey of the UK’s countryside resources, performed every 6 – 8 years (Carey *et al.* 2008). With the added insight into the macro ecological patterns of soil microbial communities, changes in soil microbial community composition can be tracked over time if the

survey continues to run. This not only allows the investigation of changes in soil microbial communities in relation to long term climatic perturbations, but also may help in the overall assessment of soil health. Tying soil function with below ground biodiversity is an important constituent of applied soil science (Bardgett *et al.* 2005). Important in regard to crop production and land use management, as well as the generation of environmental policy (Wolters *et al.* 2000), estimates of below ground microbial diversity and theories of soil microbial biogeography must be elucidated so that potential links can be made. Perhaps below ground diversity is in fact irrelevant to the overall functioning of a soil, but studies such as the ones presented here provide a preliminary understanding of below ground microbial biodiversity and how it is generated, without which there would be nothing to compare large scale functional assessments of soil with. Metagenomic approaches will eventually shed more light upon the overall functioning of different soil types (if indeed there are differences in functioning) and thus when studies of biogeography and soil function are combined, a clearer picture of important soil types and the microbial communities responsible for their functions can be made. With the vast amount of anthropogenic manipulation of edaphic conditions of soils worldwide (Harris *et al.* 2005), soil conservation may be important for modulating and maintaining the biogeochemical processes performed by soils essential to the sustainable future of life on the planet. Tying functional measures with estimates of below ground biodiversity, and in particular elucidating if particular microbial taxa are important for specific soil functions, will allow a greater understanding of the functional resilience and resistance to anthropogenic or other external perturbations of the soil environment. A clear example of this can be seen in the case of low-pH ammonia-oxidizing archaea (Gubry-Rangin *et al.* 2011). Within soils of low pH, archaea seem to be solely responsible for ammonia oxidation, thus definitively linking biodiversity to essential soil function. The same may be true for other properties of interest to the commercial sectors which utilise soil products for industrial gain such as production of antibiotics, fungicides, and enzymes. Understanding large scale biogeographic processes is a primary step in linking soil function to biodiversity, without which overall understanding of how ecosystems operate at the large and local scale will be hindered.

8.4 Methodological Limitations

All assessments of microbial biodiversity and community structure are limited in some way, and a general understanding of these limitations provides a more robust interpretation of results. Within nucleic acid based assessments, the initial bias introduced is that of the nucleic acid extraction procedure used (Feinstein *et al.* 2009). Here, rigorous scientific testing of the chosen method was carried out prior to the generation of a large scale soil nucleic acid archive, yet biases may still have been introduced. Fungi have a markedly different internal cellular structure to soil bacteria (Lipke *et al.* 1998); the use of a physical and chemical lysis step seems to be the best available method to overcome this problem (Yeates *et al.* 1998).

PCR biases introduced during the amplification of the taxonomic marker gene of choice are generally unavoidable as genes in high abundance will be preferentially amplified over those in lower abundance (Luders & Friedrich, 2003). The future aim of molecular based assessments of microbial communities is to use methods which do not require PCR. However, at present, methods such as cell sorting (Davy & Kell, 1996) and single cell sequencing (Salman *et al.* 2011) are not capable of giving an overall representation of community structure at the landscape scale. Primer biases also factor into this problem as, although the taxonomic marker genes are present within all of the microbial lineages under study, conserved primer binding sites may vary across taxa, and even though degenerate primers can be used, nonspecific amplification occurs at unavoidably higher rates (Suzuki *et al.* 1996).

Along with t-RFLP, clone library analysis and next generation sequencing are also PCR reliant, and thus fall foul of these problems. Further biases may be introduced during the sequencing and analytical stages (Kunin *et al.* 2010). Chimeric sequences and homopolymer regions all cause problems, although there are analytical procedures available to reduce the error rate inherent to sequencing (Haas *et al.* 2011; Quince *et al.* 2009). Accurate sequence alignments and the construction of OTUs via de-replication procedures are relatively straightforward when using a comparatively small number of sequences, as is the case with traditional clone library analysis. However, with the number of reads generated by next generation sequencing, short-cuts are in some cases unavoidable to analyse data within a realistic time frame. New

alignment, clustering, and taxonomy assignment methods, such as those implemented in the Pynast, UBLAST, USEARCH, and Muscle algorithms (Caporaso *et al.* 2010; Edgar, 2004; Edgar, 2010) increase the speed of calculations at the expense of accuracy, and thus a critical eye must be cast over sequence alignments and generated OTUs calculated by these methods. Unfortunately, due to the computational power required to perform full pairwise sequence alignments, alternative options are limited for use with high throughput sequencing. The same goes for phylogenetic tree building procedures such as bootstrapping. Unless stated in the text, bootstrapping was avoided due to the large amount of time taken to construct phylogenetic trees from multiple thousands of sequences. Therefore generally only a rough approximation of phylogenetic placement can be generated which, although not ideal, gives an overall insight into taxonomic descent of specific lineages of interest.

As noted in Chapter 4, the between OTU sequence variability of different taxonomic marker genes is not constant, and the sequence variability of different regions of the same taxonomically informative gene is also different. This then highlights the point of matching the variability of the region of choice with the scale and grain of the study, as previously discussed (Chapter 5). Furthermore, comparing patterns of biodiversity between microbial kingdoms using different taxonomic marker genes as ecological community descriptors potentially leads to inaccurate portrayals of the environmental and spatial effects acting upon these communities. Only rough approximations of the taxonomic resolution inherent to indirect community profiling methods (such as t-RFLP and DGGE) can be made. Therefore not only is an understanding of the taxonomic resolution inherent to a particular community profiling method needed, but also an understanding of the resolution provided by each gene used in conjunction with the plethora of molecular based community profiling methods available to the modern environmental microbiologist is vital.

The reliability of phylogenetic insight generated by using a single taxonomically informative gene has been questioned by some (Fox *et al.* 1992), and generally no consensus has been reached. Although at present this is the best option available to molecular microbiologists, the variability in single regions of a gene is probably not representative of the variability across a whole genome. The use

of multiple coding genes is generally unreasonable for studies such as those presented here, as using a second or third gene would double the time and financial expense. Use of single copy genes such as *rpoB* has been proposed as a complement to 16S rDNA analysis to provide species or subspecies level classifications, but this has yet to be widely adopted in environmental studies of complex communities (Case *et al.* 2007). True metagenomic studies, such as those proposed within the Terragenome project (International Soil Metagenome Sequencing Consortium, 2012), will eventually supersede single gene assessments of microbial community structure but the technical and analytical complexity of performing this type of study upon even a single soil sample is outside the reach of today's scientists. Until we have the ability to sequence the genome of every organism within every soil accurately and rapidly we must do the best with the tools we have, and analyse results with a critical eye and without a misguided assumption of infallibility.

8.5 Future Work

The next logical stage to this study is to attempt to relate soil function to soil microbial biodiversity and biogeography. The fact that different meta populations of soil bacterial and fungal taxa exist within soils of different physicochemical composition suggests that biological functioning of soil may not be the same in all soil ecosystems. This hypothesis allows for the investigation of more applied aspects of soil microbial ecology, principally the idea of functional redundancy (as in Yin *et al.* 2000). As, within this study, all taxa are not found everywhere, does this lead one to expect that this may be the case for functional genes also? Although there is evidence for functional redundancy in specific systems such as in carbon mineralisation (Andren & Balandreau, 1999; Rousk *et al.* 2009), specific associations have been observed between soil microbial community structure and function (Griffiths *et al.* 2000). The use of metagenomic approaches, sequencing randomly amplified fragments of the total nucleic acid component of specific soil systems, is currently underway to analyse functional diversity across a natural pH gradient. Next generation sequencing analysis of other soil microbial domains, such as protists and micro arthropods, is also being performed in order to compare environmental and spatial patterns in organisms with more complex cellular organisation. Results of these studies may truly transcend the barrier between

the microbial world and macro-bial world in an effort to elucidate a unifying theory of biogeography for organisms of a range of sizes and evolutionary descent.

Of particular interest is the effect of fine scale biogeographic patterns across short environmental and spatial gradients. Having determined macro–ecological patterns relating environmental variability and spatial separation within the bacterial and fungal kingdoms at the landscape scale, the next step is to apply hypotheses derived from these studies to areas of similar ecological characteristics and at smaller spatial scales. What is meant by this is that, as pH has been shown to be an important factor in community assembly in both bacterial and fungal communities (Chapters 3-7), and soils of the same pH exhibit different patterns of biodiversity close to the species level (Chapters 4 and 6), the factors affecting soil microbial biodiversity at the local scale may be somewhat different from those at the landscape and global scale. Within the two kingdoms spatial autocorrelation is present over short spatial gradients (Chapter 7; King *et al.* 2010; Lilleskov *et al.* 2004; Nunan *et al.* 2003). At the local scale (i.e. at centimetre resolution across a field), spatial separation may be more important in terms of structuring soil microbial communities than the environment as edaphic variability is small over short distances. Ideally areas of low and high pH soils within close proximity to each other, from non-anthropogenically manipulated systems, could be examined in a continuous, high resolution, sampling design to address questions posed in relation to spatial scaling and macro-ecological patterns. Furthermore, analysis of microbial biogeography within the soil sub matrix across the μm to mm scale would also be interesting. Comparing the community dissimilarity, or even (in view of the low number of organisms which inhabit single pore spaces) full genome sequencing of microbes separated by small scale environmental barriers (air pockets or mineral grains) could provide an insight into mechanisms of speciation. However, at present this would be technically challenging.

Patterns of microbial succession and colonisation are also of interest as they relate directly to soil formation, aging, and function. Low and high pH soils could be collected from the environment, sterilised via fumigation and transplanted into the contrasting soil habitat (as well as sterilised control soils replaced into

their original environment). Continuous, high resolution sampling of these soils, using next generation sequencing strategies or targeted quantitative PCR (e.g. for the group 1 and 6 Acidobacteria), would determine whether successional patterns are the same within both systems, and whether the climax community of each transferred soil type is the same as the natural climax community found within soils of low and high pH naturally occurring in the environment. Furthermore, interpolation of ordination scores across the soils would create an animated map of successional patterns through space and time within the low and high pH systems under study. This may provide useful insights into keystone organisms required for successful colonisation of soil, if any are present, and allow an insight into soil resilience to perturbation. Coupled with this, aerial samples (as in Bowers *et al.* 2011) could be taken a few cm above the soil surface and the community structure of these could then be studied in a similar manner to that of the soil itself, producing a second animated map so that aerial as well as substrate dispersal of soil microorganisms could be studied.

As future long term climatic perturbations have been forecast (Bakkebes *et al.* 2002), and environmental variability has been shown to be important for determining soil bacterial and fungal community structure (Chapter 3 – 6), the resistance and resilience of soil communities to this forecasted climate change is important in understanding the resultant changes in below ground diversity. This is perhaps more important in relation to how soil function may change with environmental change and, if it does, how resistant soil function is to climatic variability and whether soil function can be returned to present levels post perturbation. Further work is required to understand changes to soil physicochemical composition in relation to predicted changes in climatic conditions. However, through the work carried out in this thesis, preliminary predictions relating to the resulting microbial communities within these altered soils could be made. Unfortunately, little is known about how landscape scale changes in soil physicochemical properties may affect below ground communities within single soil types. It cannot be assumed that they will mirror the communities found in soils with similar physicochemical conditions to the altered soil habitat. Translocation studies would help to predict the potential changes in below ground communities and function if soils from one area of the

globe were transplanted to another area of the globe which exhibited similar climatic conditions to those forecasted as a result of long term climate change. For example, a section of low pH peat bog from Scotland could be transplanted into a Mediterranean soil, then the microbial community structure, composition, and meta transcriptome could be analysed over time to give an indication of how the microbial community structure and function of this soil type may change due to climate change.

8.6 Hypothetical Scenarios: The Perfect Study

At the time of this study, the spatial and taxonomic resolutions of the landscape scale surveys of bacterial and fungal communities were at the boundary of possibility. That is, higher spatial and taxonomic resolutions would be technically and financially hard to achieve due to the larger number of samples required, and computational expense needed, to generate and analyse microbial community data at high spatial resolutions across a landscape. Ideally, to fully analyse biogeographic patterns at multiple spatial scales at once, high resolution sampling designs at the metre or even centimetre scale would be needed across an entire landscape such as the UK. If used in combination with high resolution, taxonomically informative community profiling techniques such as next generation sequencing, this would, in essence, not only make it possible to elucidate broad scale patterns in biogeography, but at the same time, local scale patterns of biogeography could be linked with landscape scale patterns in a unifying theory of microbial biogeography at all spatial scales.

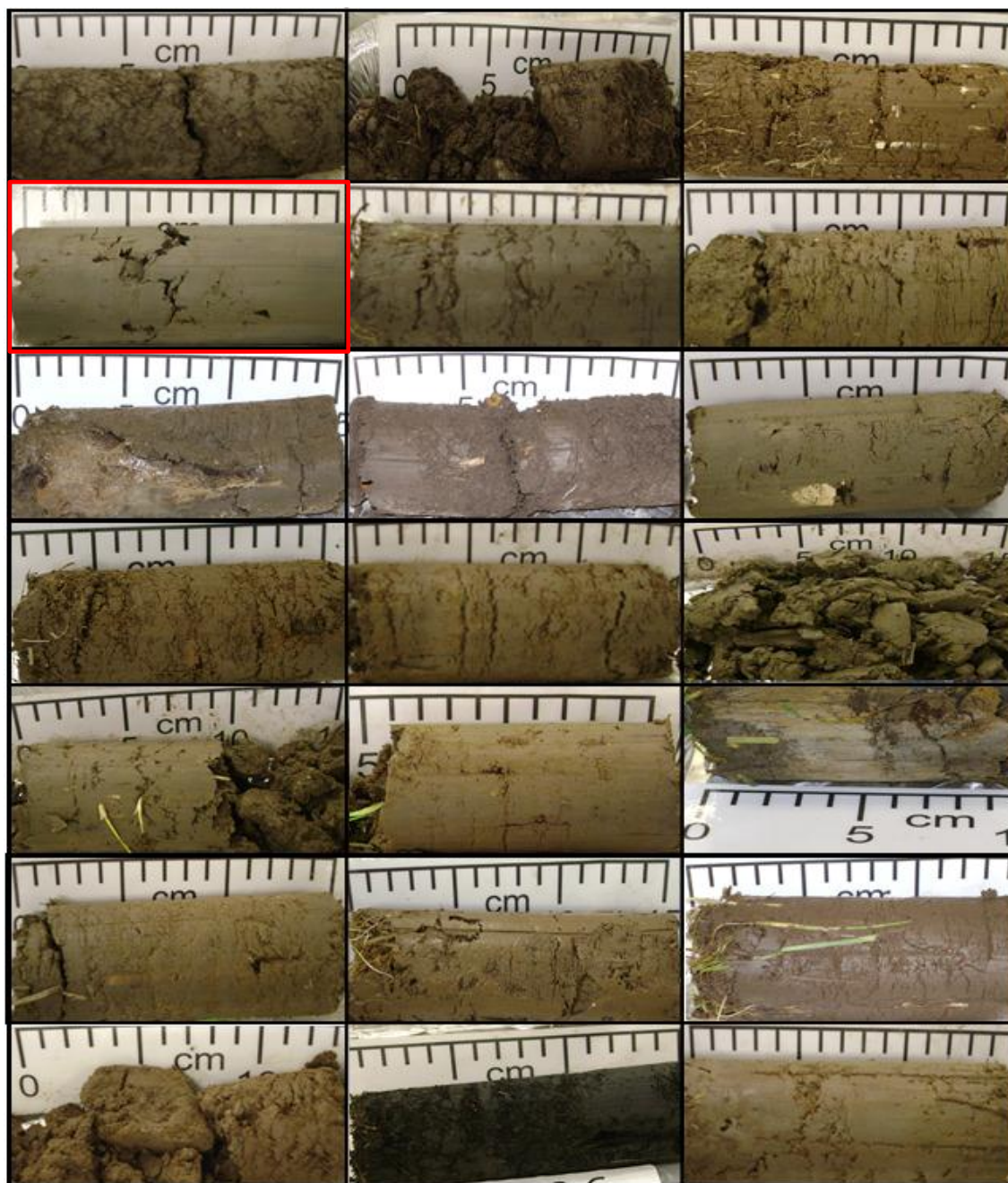
8.7 Final Conclusions

Landscape scale patterns of soil bacterial and fungal biodiversity are related to environmental variability. However, variation in bacterial populations seems to be more related to differences in environmental conditions than is variation in fungal populations. Although similarities between the biogeography of the two kingdoms were seen in relation to variability in relation to environmental and spatial factors, different environmental characteristics were identified as the most important determinants of bacterial and fungal community structure. Bacterial populations are structured in relation to a number of co-correlating

environmental conditions as are soil fungal populations, but the former show strong relationships with pH, the latter with the above ground plant communities, although both factors were found to be highly significant and important community determinants for both kingdoms. It may be hypothesised that prokaryotic lineages are more susceptible to local environmental conditions than eukaryotic lineages due to differences in individual size and dispersal ability. Nevertheless, both kingdoms can be seen to conform to many of the traditional hypotheses derived from observations of macro-bial biogeography. Phylum level patterns show distinct relationships with environmental gradients, but when OTUs were classified closer to the species level, areas of similar physicochemical properties exhibited different patterns of biodiversity as is seen in macro-bial populations. Therefore, it can be hypothesised that, although the “environment selects” it may not be the only factor affecting the biogeography of these kingdoms of microbial life as often proposed (O'Malley, 2007). The taxonomic level at which patterns in biodiversity are analysed will ultimately affect the conclusions drawn from such studies, and to put this into context, examining phylum level patterns across a landscape within the microbial world is akin to analysing the biogeography of the Arthropoda across the UK. Take any soil in the world and you will find arthropods, from the Arctic (Bale *et al.* 2007) to the tropics (Giller, 1996). Without analysis at higher levels of taxonomic organisation, one would be forgiven for concluding that arthropods are everywhere but the environment selects how many there are at a given location, and thus they do not conform to laws governing the distribution of easily observable macro organisms such as mammalian species. Of course we know that individual members of the Arthropoda have defined spatial and environmental niches which are controlled by the given range of specific organisms, due to dispersal limitations, competition, predation and so on (Adler & Dudley, 2008) and thus they do in fact conform to the generally accepted biogeographic principle. When put like this, it becomes apparent just how little we know about soil microbial life in comparison to that of macro organisms, and that microbial biogeography is a very young discipline within which we have only just started to explore the most basic and overarching patterns of biodiversity. In many cases, these patterns are complex and the general understanding of them is still vague. As technology advances and allows a more detailed exploration into the biogeography of microbial species rather than

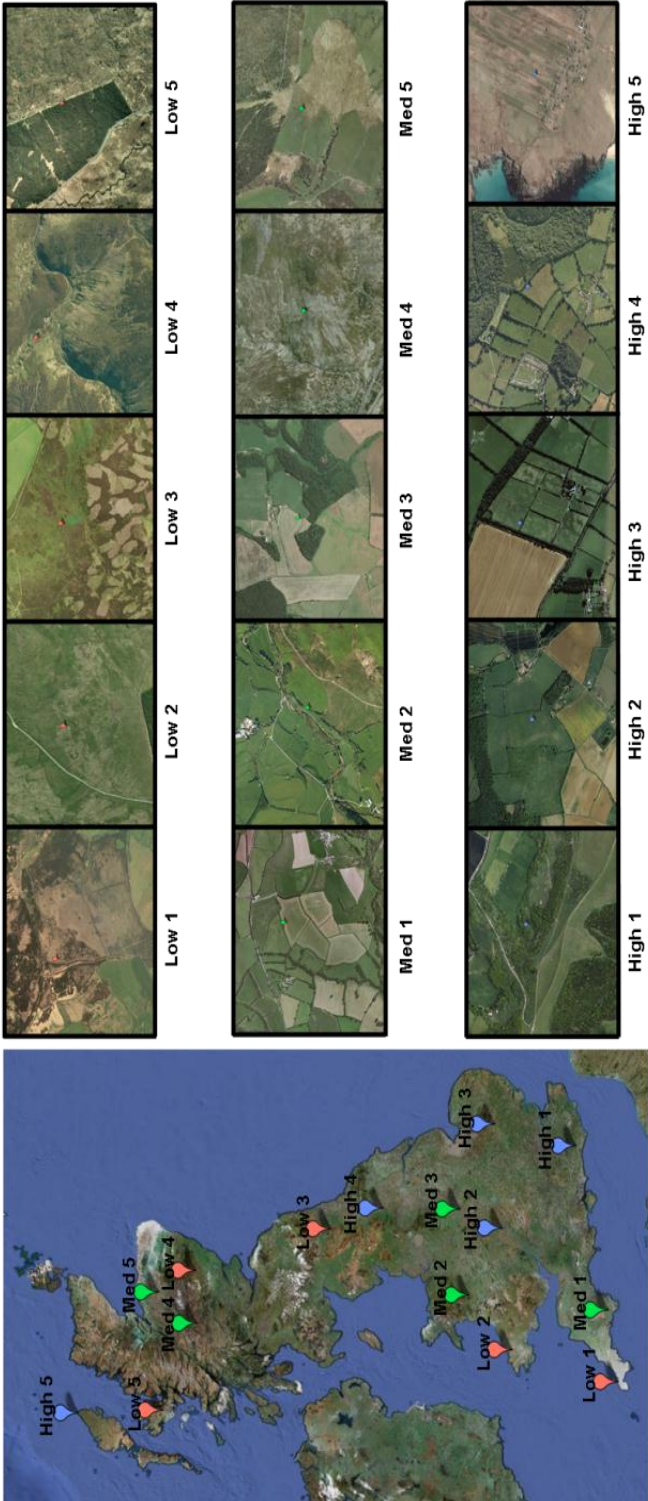
phylum and class level assessments, these patterns are likely to change and become even more complex. I personally believe that, as a result of the work detailed above, we will find that true species level assessments of soil microbial biogeography and biodiversity describe the familiar patterns seen across the macro-bial biosphere, and that from *E. coli* to elephants, no great divide exists between the ecological principles of the very small or the very large, just our perception.

Appendix A

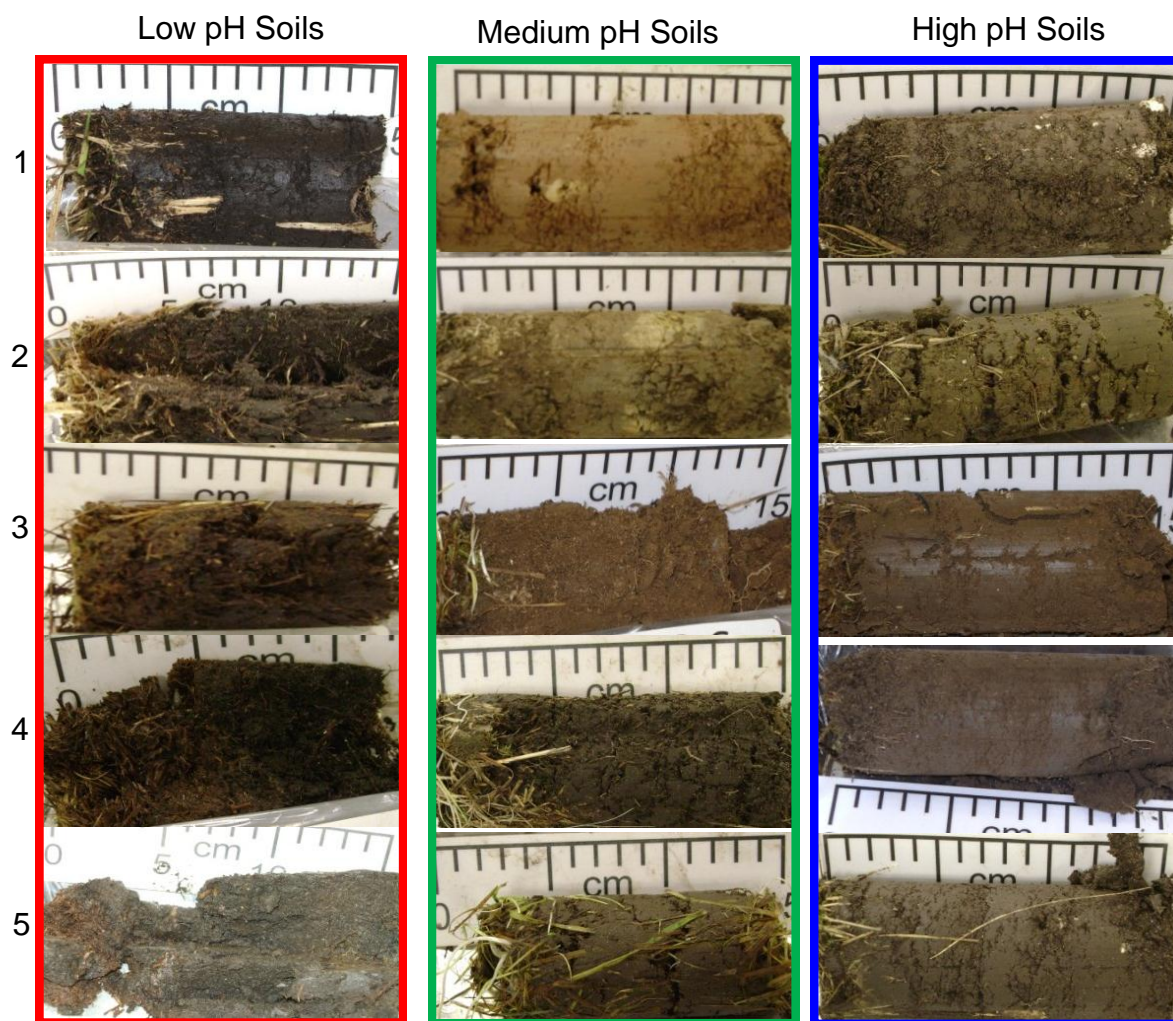


Appendix A: Soil samples with high clay contents (as defined by texture analysis) used to compare the nucleic acid extraction efficiency of the original Griffiths method and the modified nucleic extraction method using a reduction in sample mass and freeze thaw step. The soil sample from which no nucleic acid extract was obtained using either method is highlighted with a red box.

Approximate Locations of the 15 Samples Used for Pyrosequencing
and High Resolution Photographs of the Sample Sites



Appendix B: The locations of the 15 subsamples (5 low pH, 5 medium pH, and 5 high pH) used for pyrosequencing. Map locations show sample locations within a tolerance of 5 KM and high resolution satellite images show sample position within a tolerance of ~ 10 m. Satellite imagery was obtained from Google maps after sample locations were converted from eastings and northings to longitude and latitude.



Appendix B: Soil Cores used for multiple nucleic extraction test, and clone library, and pyrosequencing analysis of bacterial and fungal populations.

Appendix C

i)	C&W	FG	H&B	IG	LW	MGM	TG&H	UW
C&W	N/A							
FG	< 0.001	N/A						
H&B	0.26	< 0.001	N/A					
IG	<0.001	0.99	< 0.001	N/A				
LW	0.07	0.47	0.65	0.24	N/A			
MGM	< 0.001	0.97	< 0.001	0.71	0.88	N/A		
TG&H	0.37	< 0.001	0.99	< 0.001	0.96	0.04	N/A	
UW	< 0.001	0.84	< 0.001	0.46	0.97	0.99	0.15	N/A

ii)	C&W	FG	H&B	IG	LW	MGM	TG&H	UW
C&W	N/A							
FG	< 0.001	N/A						
H&B	< 0.001	< 0.001	N/A					
IG	0.052	0.54	< 0.001	N/A				
LW	1.00	0.20	< 0.001	0.75	N/A			
MGM	< 0.001	0.97	0.98	< 0.001	0.0015	N/A		
TG&H	0.98	0.0015	< 0.001	0.078	0.99	< 0.001	N/A	
UW	< 0.001	< 0.001	0.99	< 0.001	< 0.001	0.97	< 0.001	N/A

iii)	C&W	FG	H&B	IG	LW	MGM	TG&H	UW
C&W	N/A							
FG	< 0.001	N/A						
H&B	0.99	< 0.001	N/A					
IG	< 0.001	0.99	< 0.001	N/A				
LW	0.99	0.78	0.91	0.87	N/A			
MGM	0.29	< 0.009	0.69	< 0.001	0.39	N/A		
TG&H	0.99	0.17	0.97	0.26	0.99	0.41	N/A	
UW	0.006	< 0.001	0.04	< 0.001	0.05	0.86	0.03	N/A

Appendix C: Table of significance values (P) from Tukey's honest significance test showing pairwise comparisons of mean: i) DNA yield; ii) 260:230 nm ratios; and iii) 260:280 nm ratios in each aggregate vegetation classification. Significant differences in variable means are highlighted in grey. C&W = Crops and weeds, FG= Fertile grassland, H&B = Heath and bog, LW = lowland wooded, MGM = Moorland grass mosaics, TG&H = Tall grass and herb, UW = Upland wooded, IG = Improved Grassland.

Appendix D

i)	Clay	Clay-Loam	Loam	Peat	Sand	Silty-Loam
Clay	N/A					
Clay-Loam	0.39	N/A				
Loam	< 0.001	< 0.001	N/A			
Peat	< 0.001	0.052	0.97	N/A		
Sand	< 0.001	0.007	0.99	0.99	N/A	
Silty-Loam	0.002	0.047	0.12	0.82	0.46	N/A

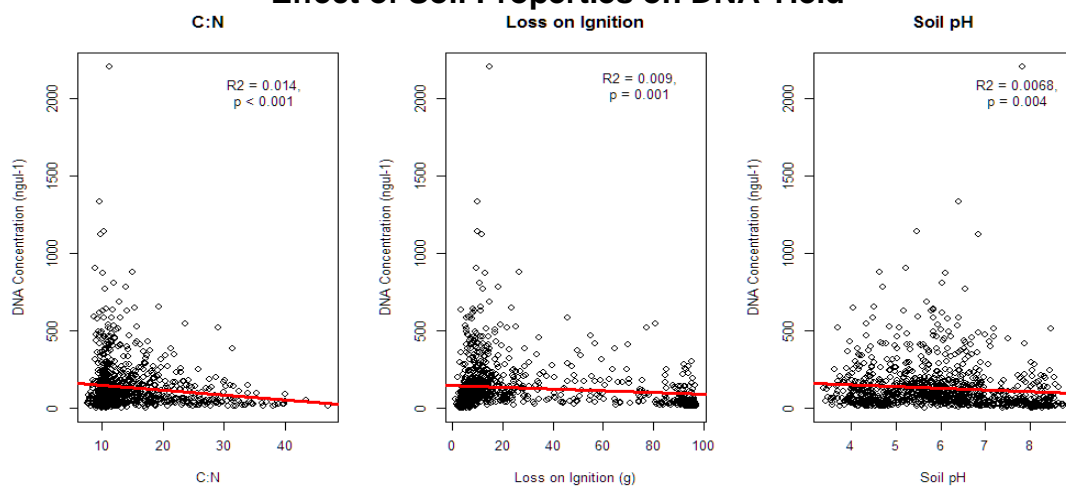
ii)	Clay	Clay-Loam	Loam	Peat	Sand	Silty-Loam
Clay	N/A					
Clay-Loam	0.99	N/A				
Loam	0.41	0.56	N/A			
Peat	< 0.001	< 0.001	< 0.001	N/A		
Sand	0.68	0.59	0.002	0.004	N/A	
Silty-Loam	0.53	0.65	0.99	< 0.001	0.015	N/A

c)	Clay	Clay-Loam	Loam	Peat	Sand	Silty-Loam
Clay	N/A					
Clay-Loam	0.14	N/A				
Loam	< 0.001	0.85	N/A			
Peat	0.99	0.08	< 0.001	N/A		
Sand	0.97	0.52	0.009	0.90	N/A	
Silty-Loam	0.16	0.99	0.66	0.09	0.59	N/A

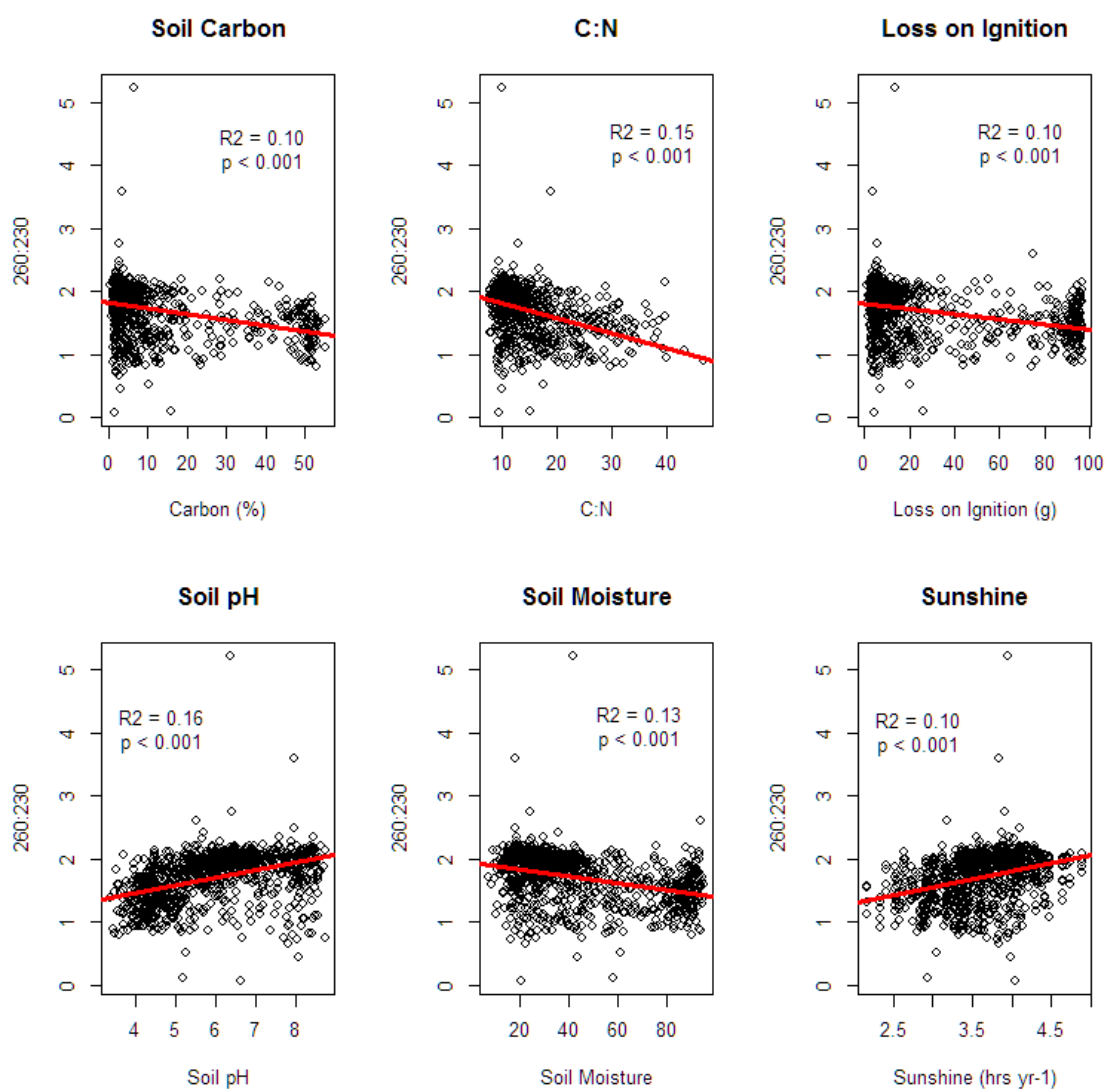
Appendix D: Table of significance values (P) from Tukey's honest significance test showing pairwise comparisons of mean: i) DNA yield; ii) 260:230 nm ratios; and iii) 260:280 nm ratios in each soil type. Significant differences in variable means are highlighted in grey.

Appendix E

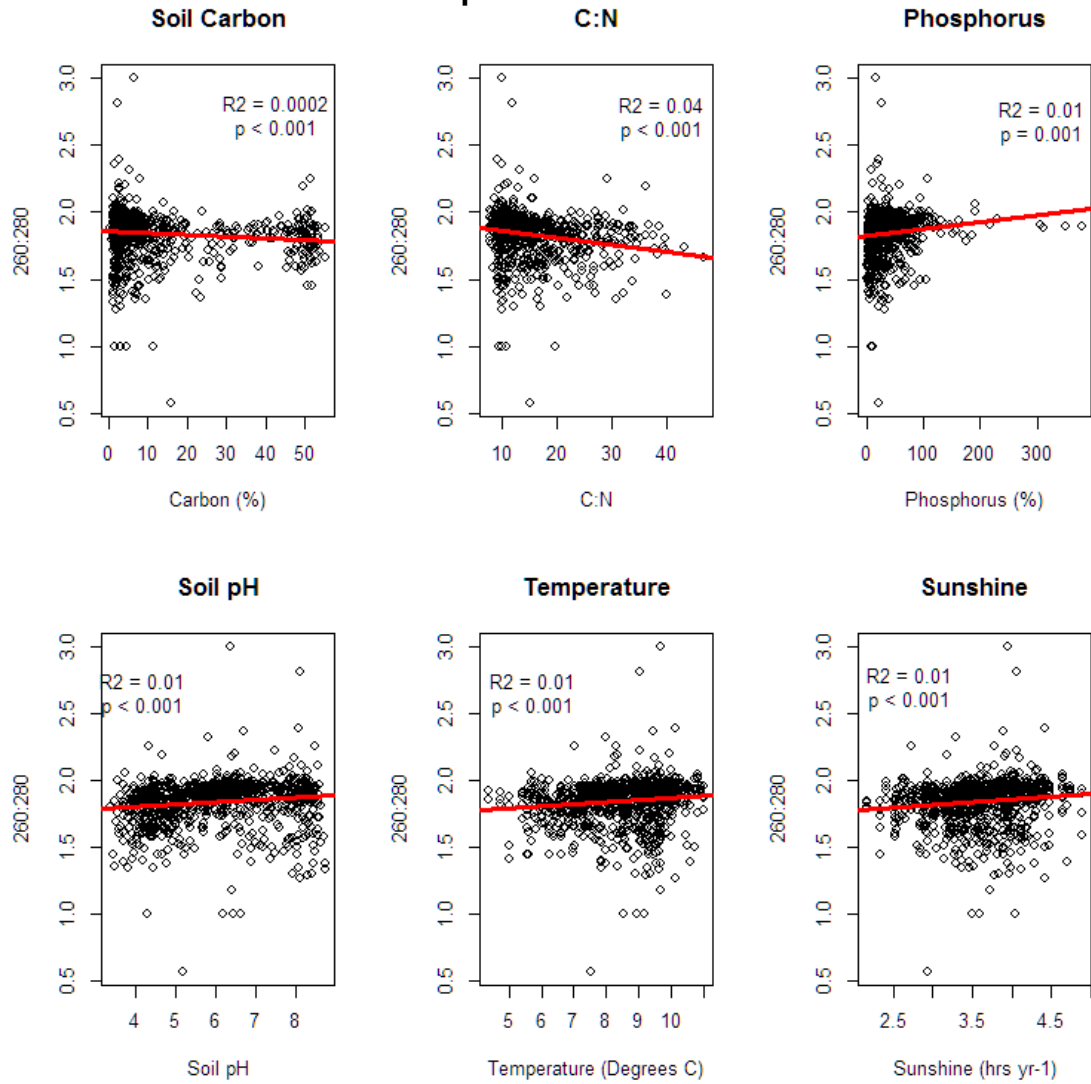
Effect of Soil Properties on DNA Yield



Effect of Soil Properties on 260 nm/230 nm Ratio

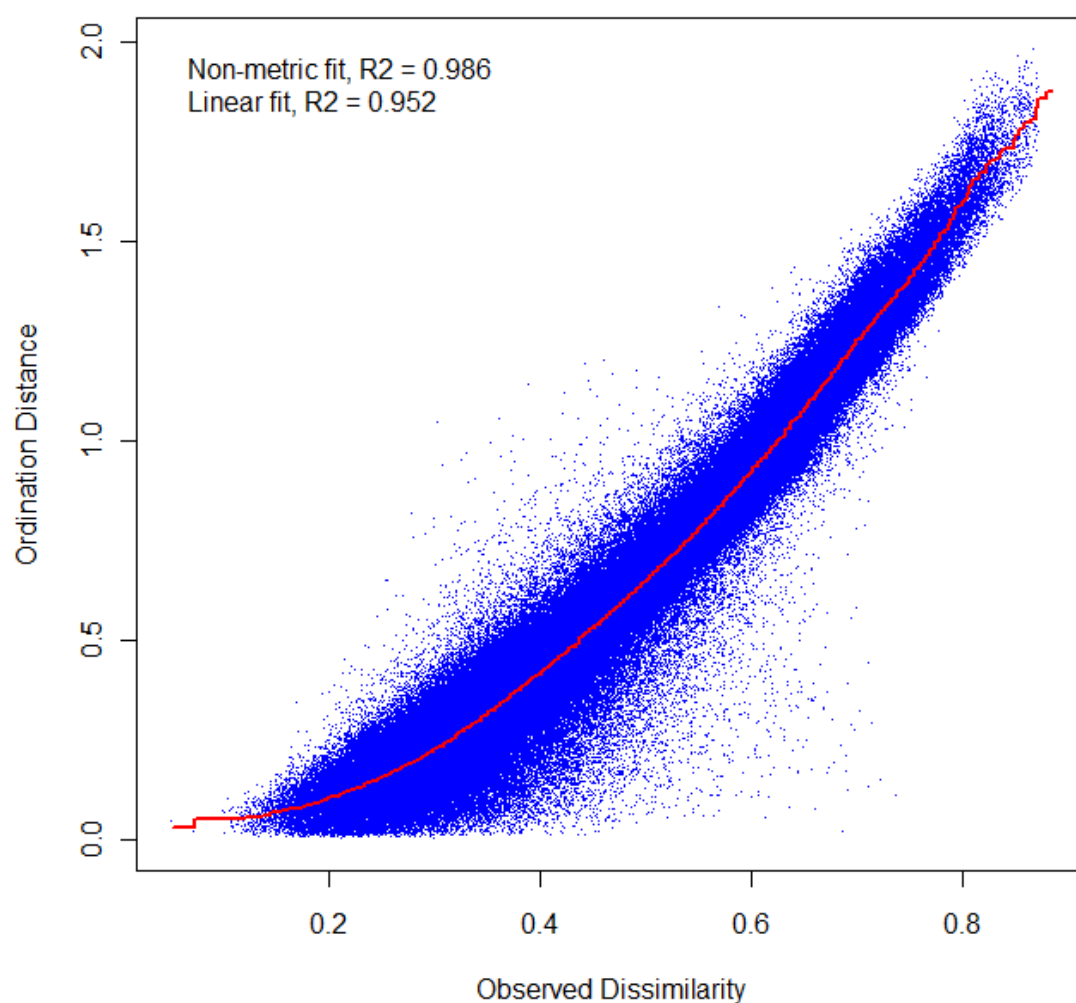


Effect of Soil Properties on 260 nm/280 nm Ratio



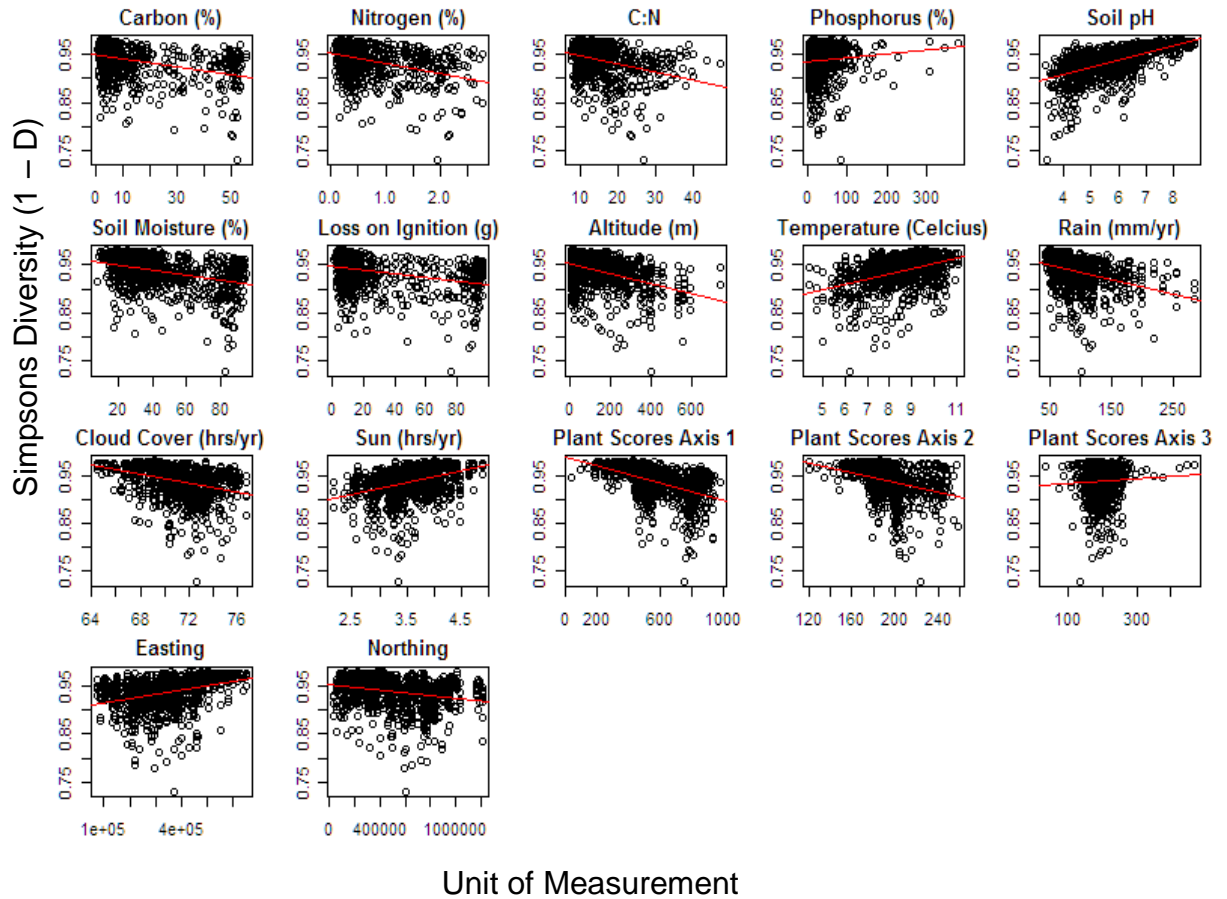
Appendix E: Linear regressions of numerical environmental variables upon nucleic acid yield, 260 nm/ 230 nm ratio and 260 nm/ 280 nm ratios.

Appendix F



Appendix F: Stress plot to show the correlation between the calculated bacterial community dissimilarity of all samples according to the t-RFLP profiles generated, and the ordination distance portrayed in the NMDS plot. The linear and non-metric fits describe a good model fit, and the low stress value (11.06) supports this.

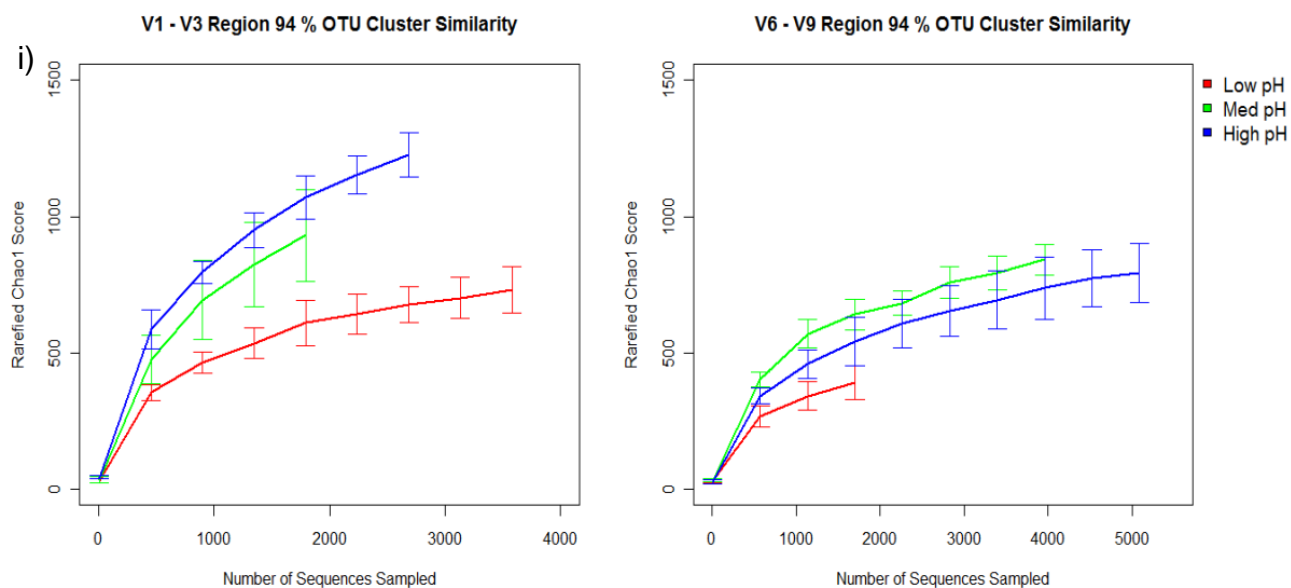
Appendix G



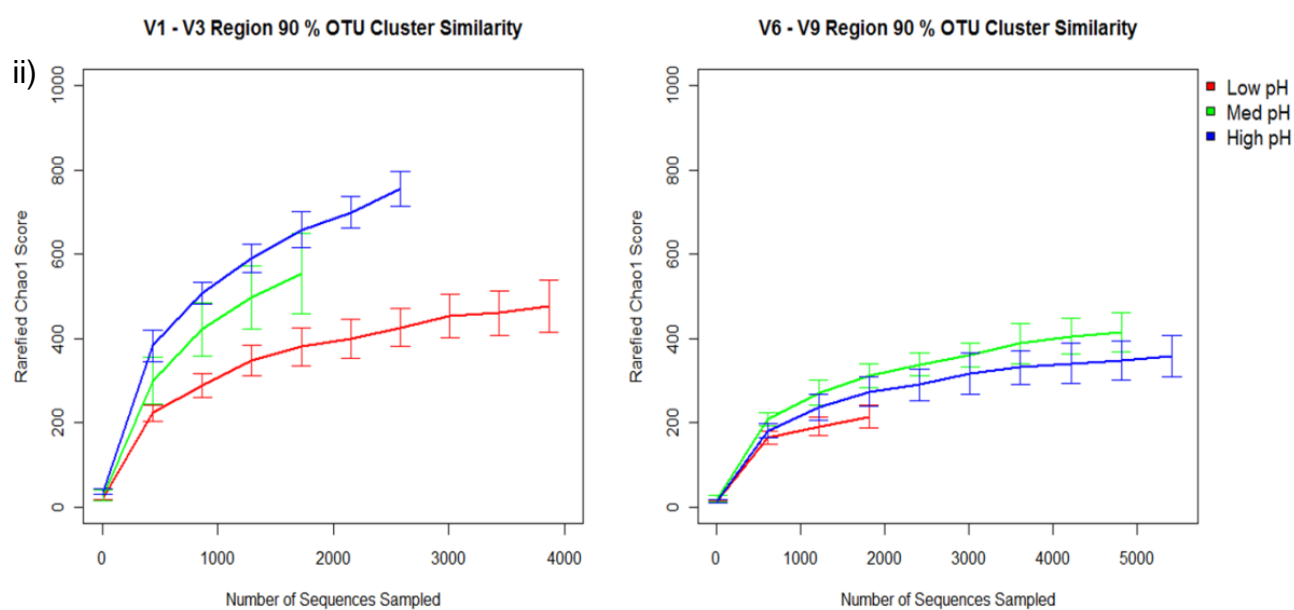
Appendix G: Correlation between all numerical environmental variables and bacterial diversity as described by Simpson's diversity index (1-D). Soil pH had the strongest correlation with diversity ($R^2 = 0.42$, $p < 0.001$) followed by the plant DCA scores axis 1 ($R^2 = 0.26$, $p < 0.001$). All other environmental variables had weak correlations with bacterial diversity; carbon ($R^2 = 0.13$, $p < 0.001$); nitrogen ($R^2 = 0.13$, $p < 0.001$); C:N ($R^2 = 0.10$, $p < 0.001$); P ($R^2 = 0.005$, $p = 0.007$); soil moisture ($R^2 = 0.15$, $p < 0.001$); loss on ignition ($R^2 = 0.12$, $p < 0.001$); altitude ($R^2 = 0.18$, $p < 0.001$); temperature ($R^2 = 0.18$, $p < 0.001$); rainfall ($R^2 = 0.14$, $p < 0.001$); cloud cover ($R^2 = 0.12$, $p < 0.001$); sunshine ($R^2 = 0.13$, $p < 0.001$); plant axis 2 ($R^2 = 0.11$, $p < 0.001$); plant axis 3 ($p > 0.01$).

Appendix H

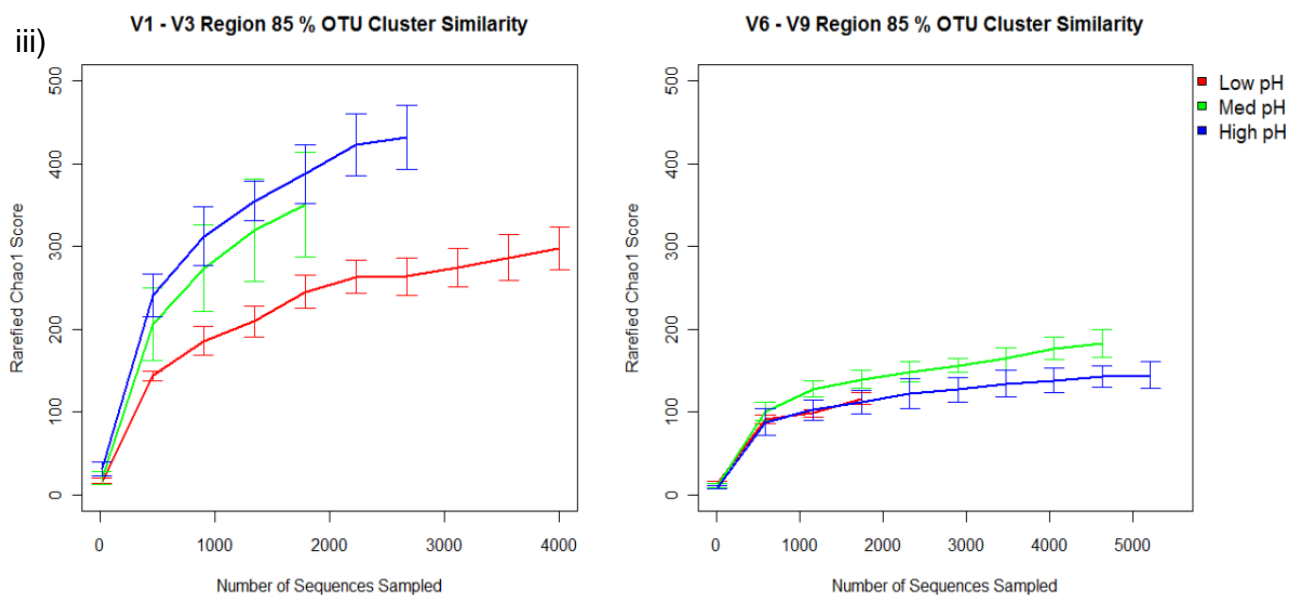
Rarefied Chao1 Scores for Bacterial Communities from Low Medium and High pH soils



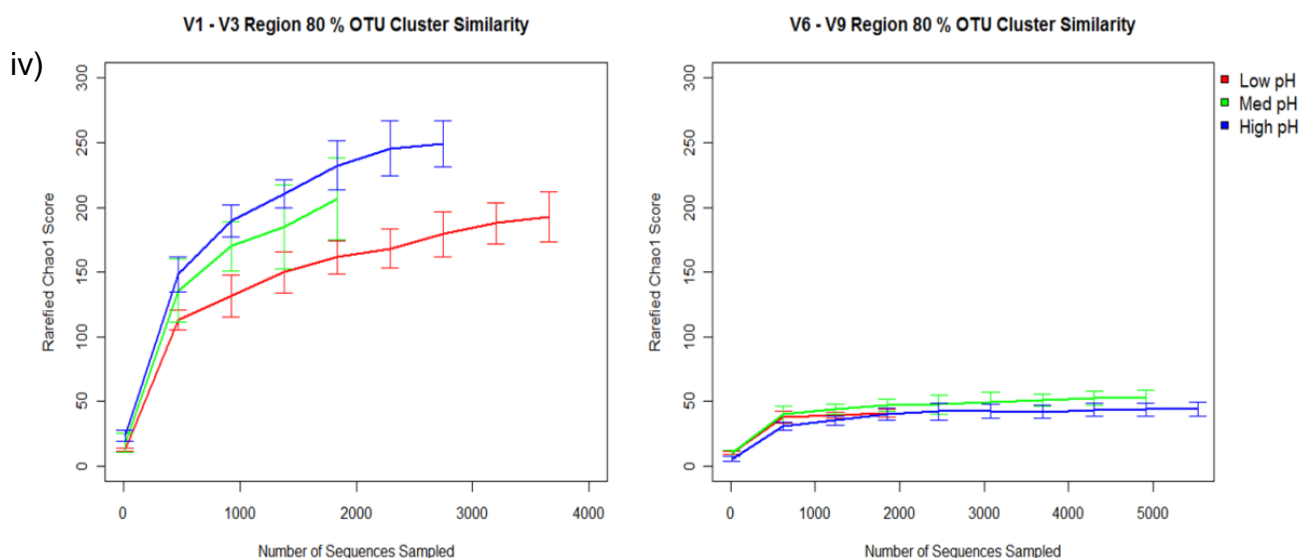
Rarefied Chao1 Scores for Bacterial Communities from Low Medium and High pH soils



Rarefied Chao1 Scores for Bacterial Communities from Low Medium and High pH soils

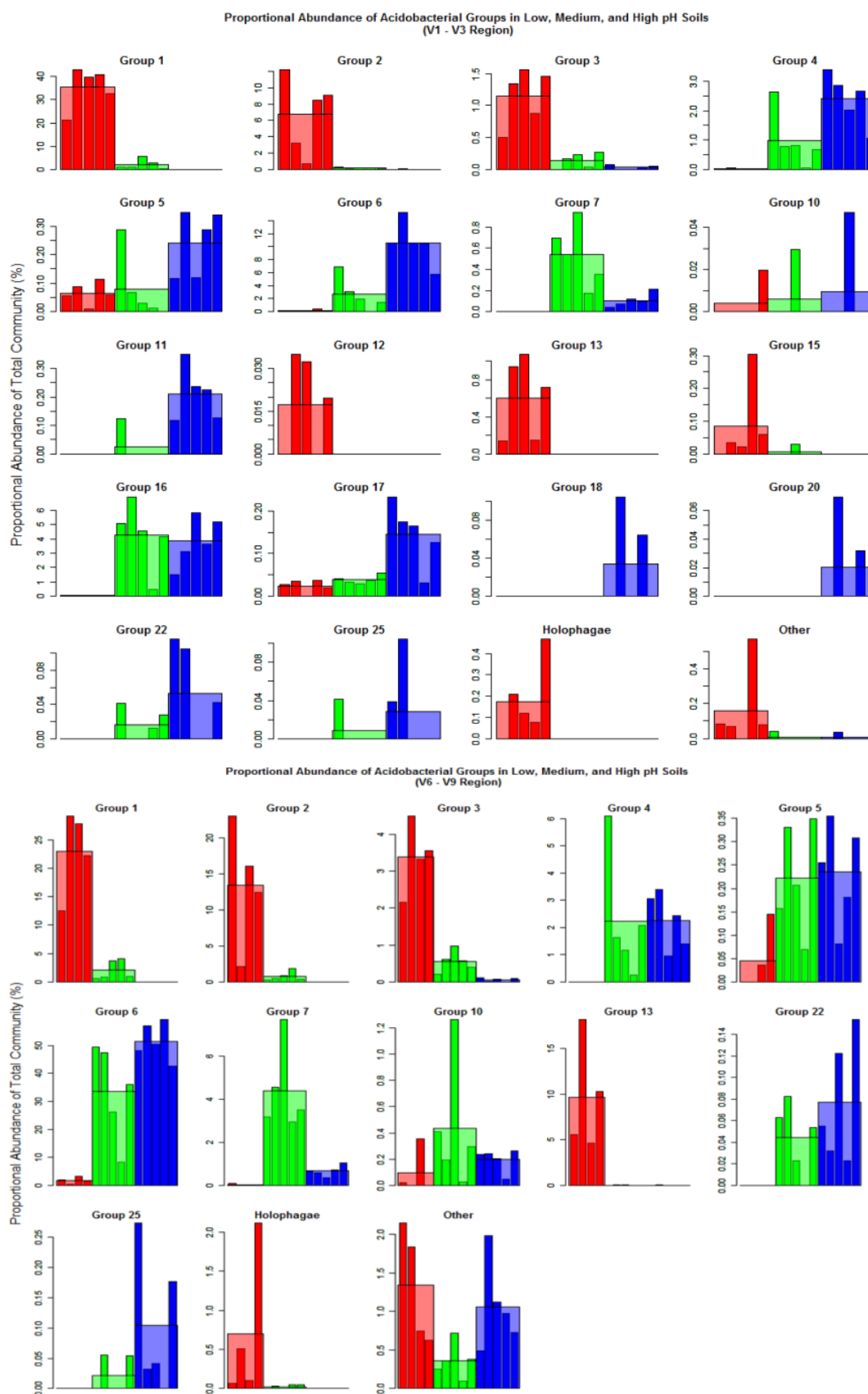


Rarefied Chao1 Scores for Bacterial Communities from Low Medium and High pH soils



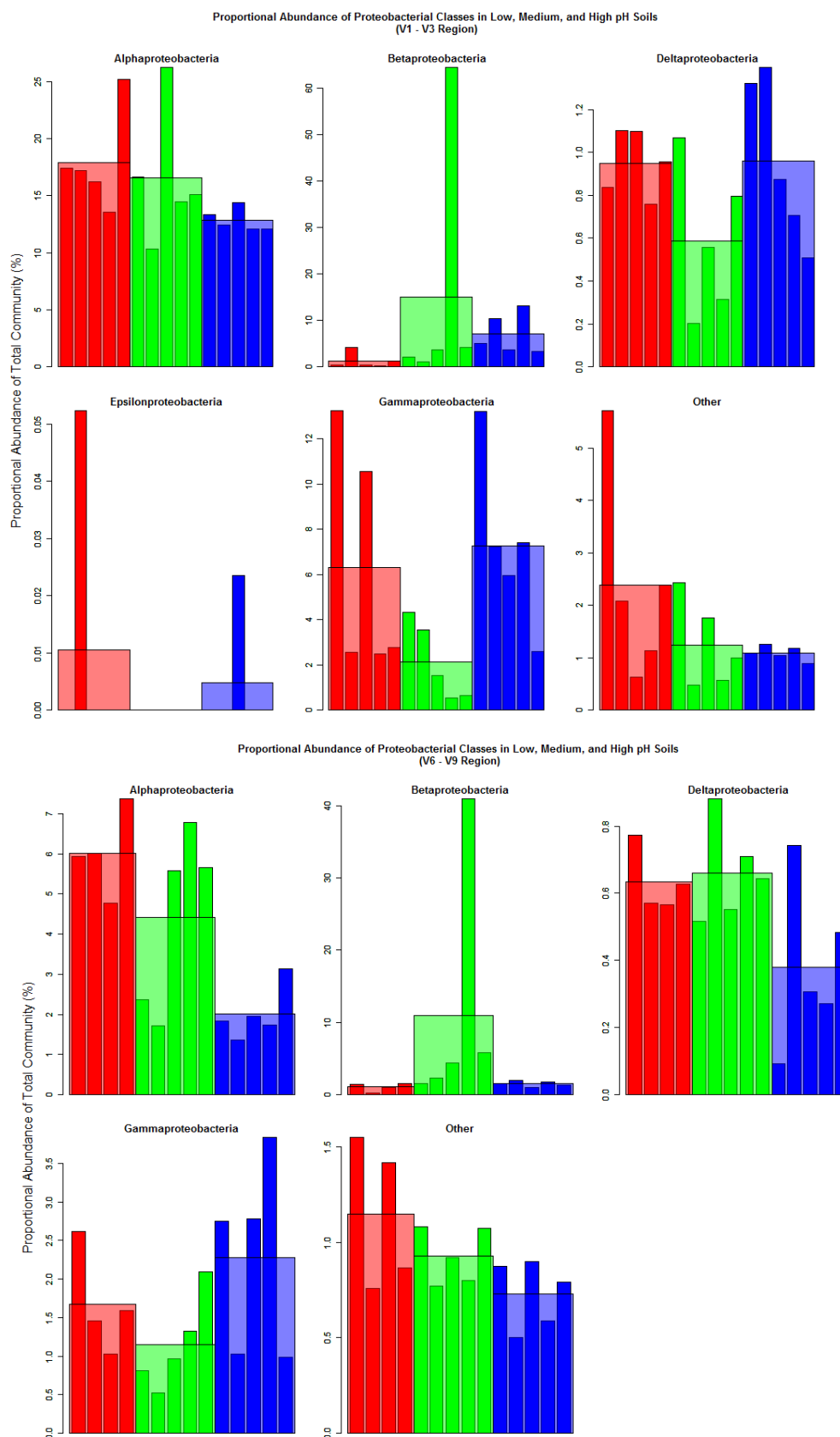
Appendix H: Mean rarefied Chao1 scores for low, medium and high pH samples from the V1 – V3 and V6-V9 regions sequences at lower levels of OTU clustering stringency. i) 94 % similarity ii) 90 % similarity iii) 85 % similarity iv) 80 % similarity. Lower diversities were reported using the V6 – V9 regions than the V1 – V3 region at all levels of OTU clustering, but low pH samples maintained a lower diversity than the medium and high pH samples using both regions of the gene. At the lowest levels of OTU clustering, medium pH samples showed a higher diversity than the high pH samples using the V6 – V9 sequences, however this trend was reversed using the V1 – V3 sequences.

Appendix I



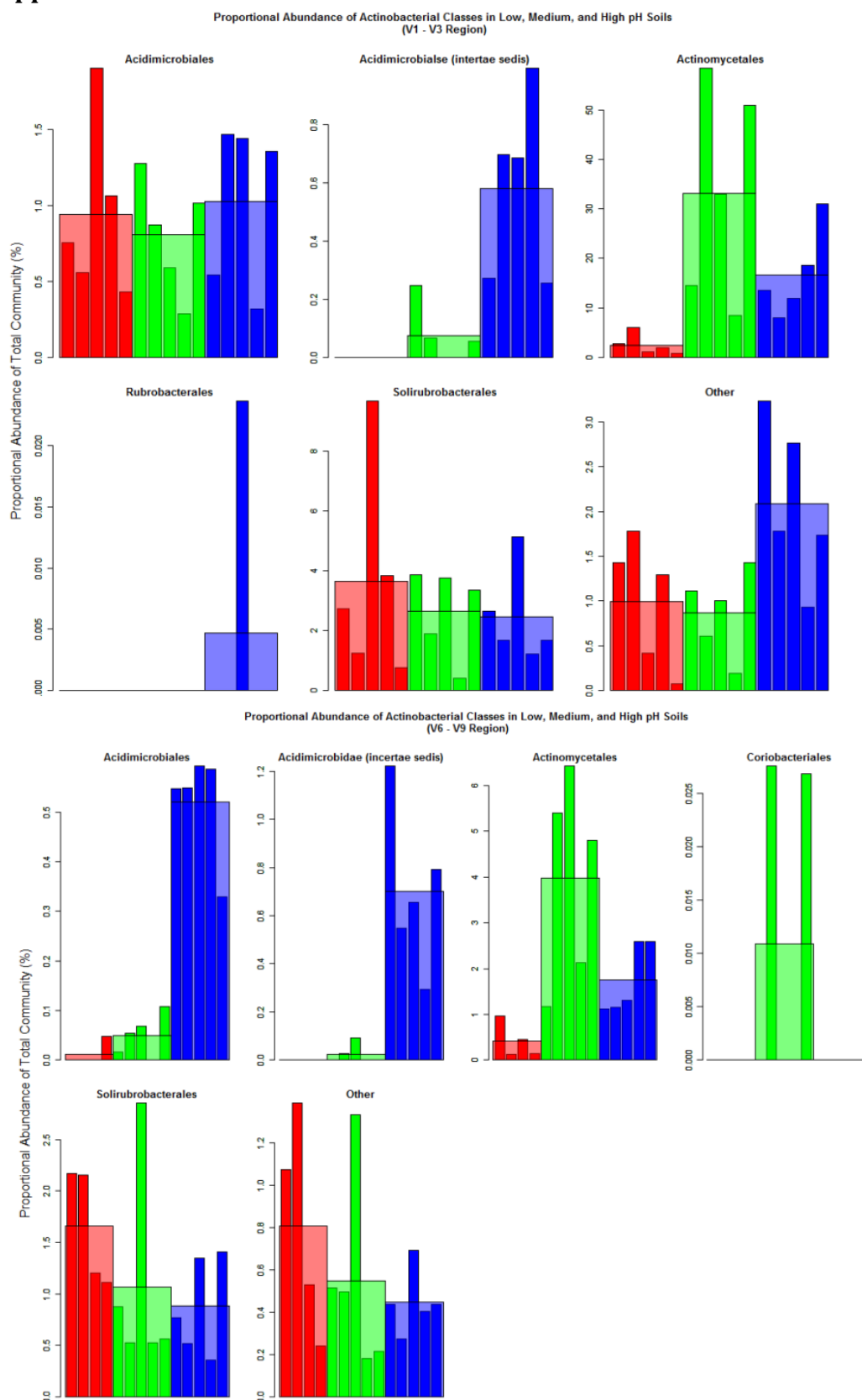
Appendix I: Proportional abundances of Acidobacterial subgroups, calculated from the sequences spanning the A) V1 – V3 regions and B) V6 – V9 region of the 16S rRNA gene. Solid bars represent individual soil samples within each pH group (Low = Red, Medium = Green, High = Blue) and translucent bars represent the mean abundance (n=5) of that taxa within a pH group. OTU clustered at 97 %.

Appendix J



Appendix J: Proportional abundances of Proteobacterial subgroups, calculated from the sequences spanning the A) V1 – V3 regions and B) V6 – V9 region of the 16S rRNA gene. Solid bars represent individual soil samples within each pH group (Low = Red, Medium = Green, High = Blue) and translucent bars represent the mean abundance (n=5) of that taxa within a pH group. OTU clustered at 97 %.

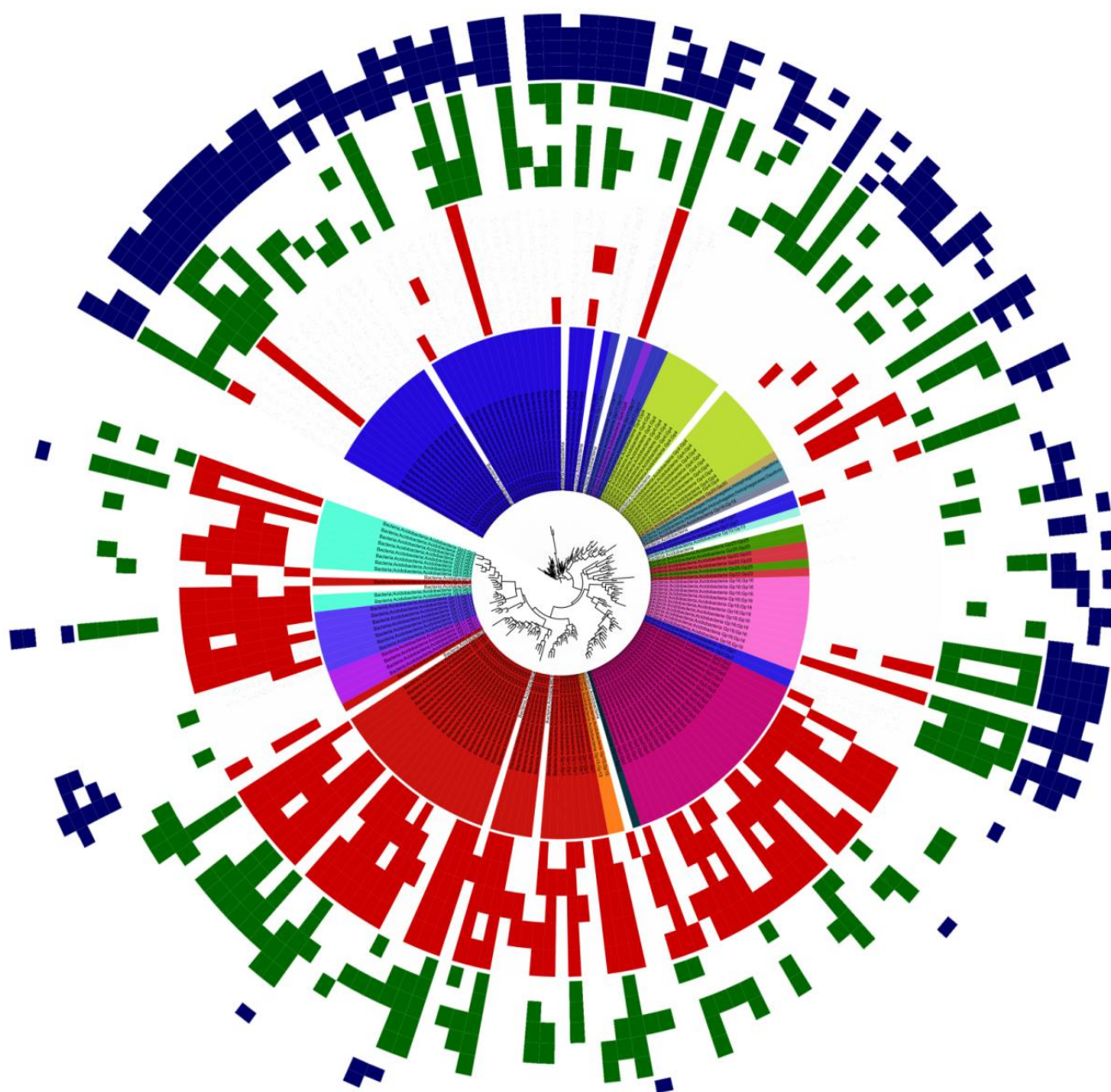
Appendix K



Appendix K: Proportional abundances of Actinobacterial subgroups, calculated from the sequences spanning the A) V1 – V3 regions and B) V6 – V9 region of the 16S rRNA gene. Solid bars represent individual soil samples within each pH group (Low = Red, Medium = Green, High = Blue) and translucent bars represent the mean abundance (n=5) of that taxa within a pH group. OTU clustered at 97 %.

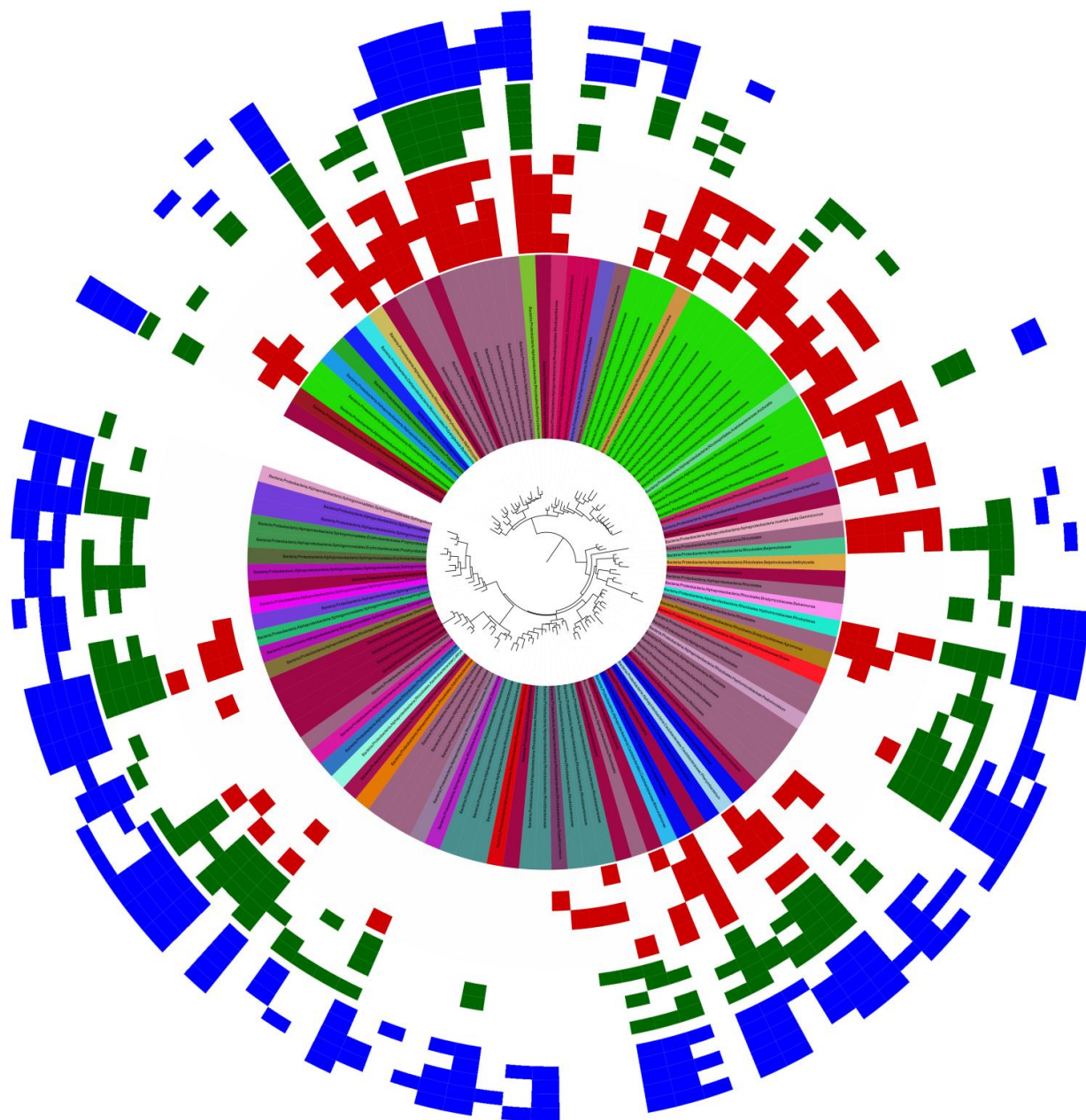
Appendix L

Phylogenetic Tree Showing the Presence or Absence of Acidobacterial OTUs Across all Samples (V1 – V3 16S rRNA gene hypervariable regions)



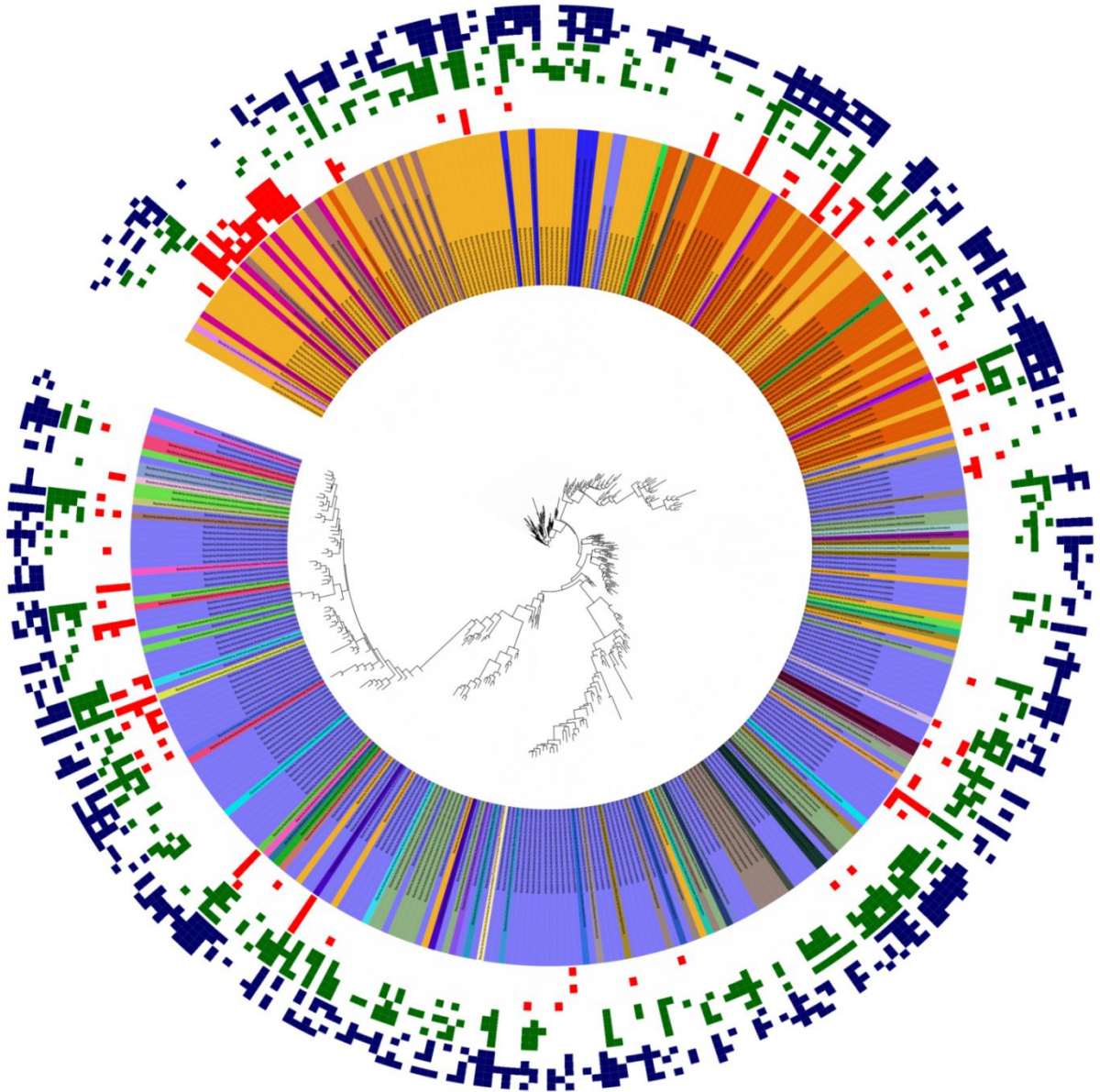
Appendix L: i) Radial heat map of Acidobacterial distribution in 5 low (red), 5 medium (green) and 5 high (blue) pH soils. Terminal nodes depict the presence or absence of each OTU within the samples. Members of the group 1 Acidobacteria can be seen to be present in most, but not all, of the low pH samples but infrequently detected in any of the high pH samples. The reverse is true for the group 6 Acidobacteria.

Phylogenetic Tree Showing the Presence or Absence of Alphaproteobacterial OTUs Across all Samples (V1 – V3 16S rRNA gene hypervariable regions)



Appendix L:ii) Phylogenetic tree and Radial heat map of Alphaproteobacterial OTU distribution in 5 low (red), 5 medium (green) and 5 high (blue) pH soil samples. Terminal nodes depict the presence or absence of each OTU within the samples. In the top proportion of the tree Rhizobial OTUs were found within most samples, however at the bottom of the tree they were predominantly found in the medium and high pH samples. This may be due to variances in species or subspecies composition detected by sequence variability (i.e. phylogenetic position) but undetected by the RDP taxonomic assignment method.

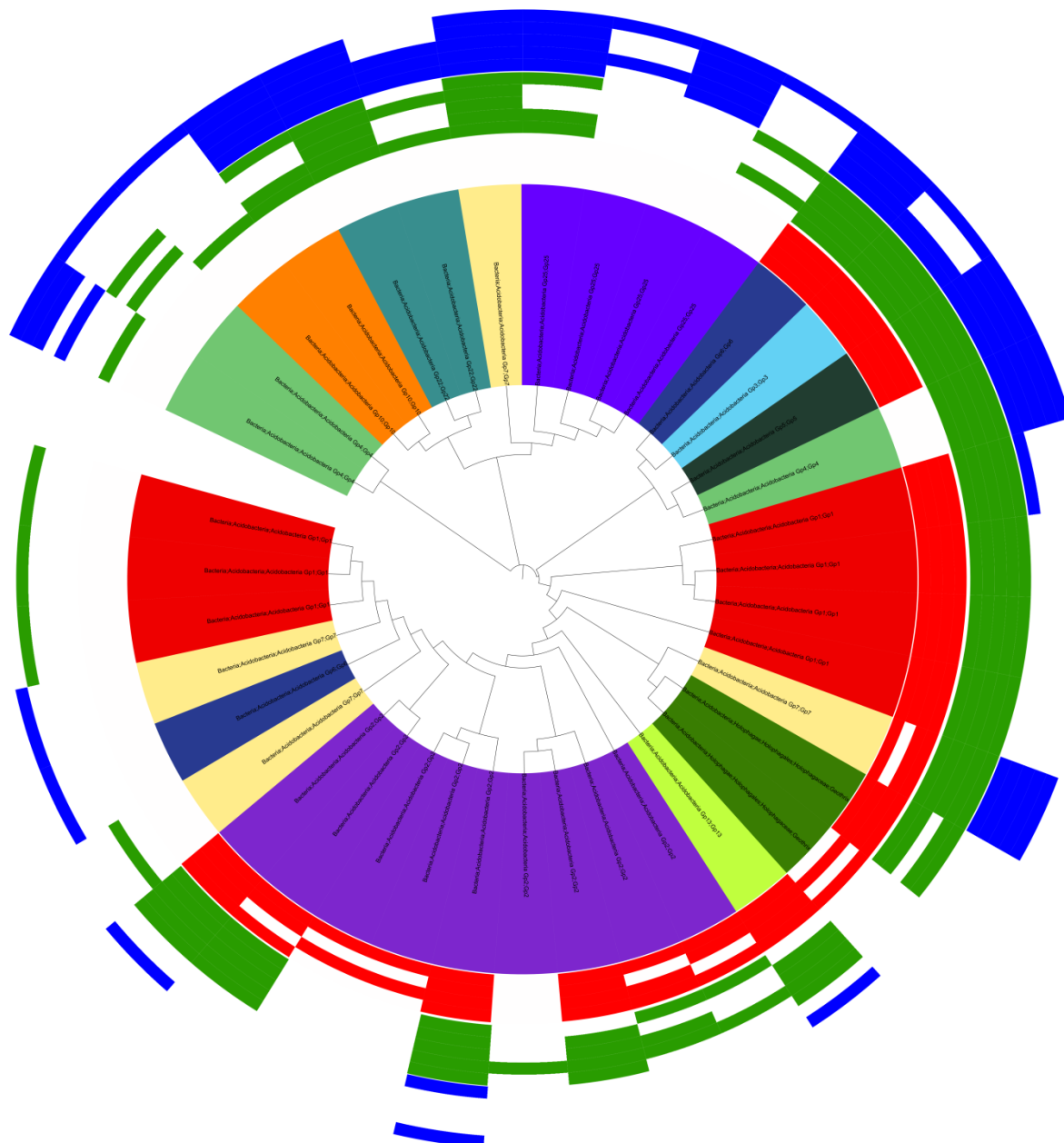
Phylogenetic Tree Showing the Presence or Absence of Actinobacterial OTUs Across all Samples (V1 – V3 16S rRNA gene hypervariable regions)



Appendix L: iii) Phylogenetic tree and radial heat maps show that Actinobacterial OTUs were predominantly found across most medium and high pH soils. The Mycroccacae or Streptomyces were occasionally found across all samples. Those predominantly defined as unique to the low pH samples were shown to belong to the aptly named Acidimicrobiales. Blue heat maps represented the high pH samples, green, the medium pH samples and red, the low pH samples.

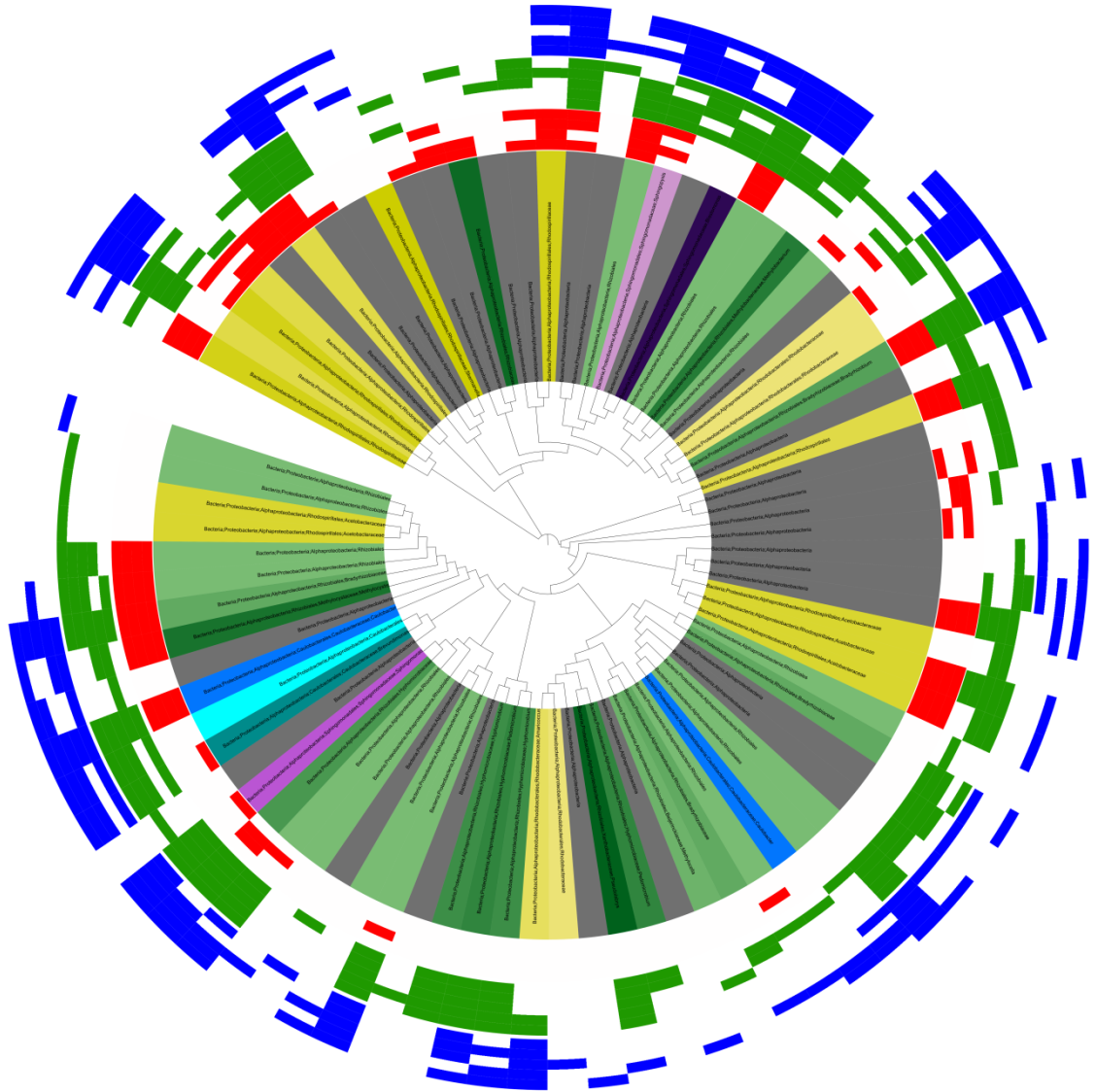
Appendix M

Phylogenetic Tree Showing the Presence or Absence of Acidobacterial OTUs Across all Samples (V6 – V9 16S rRNA gene hypervariable regions)



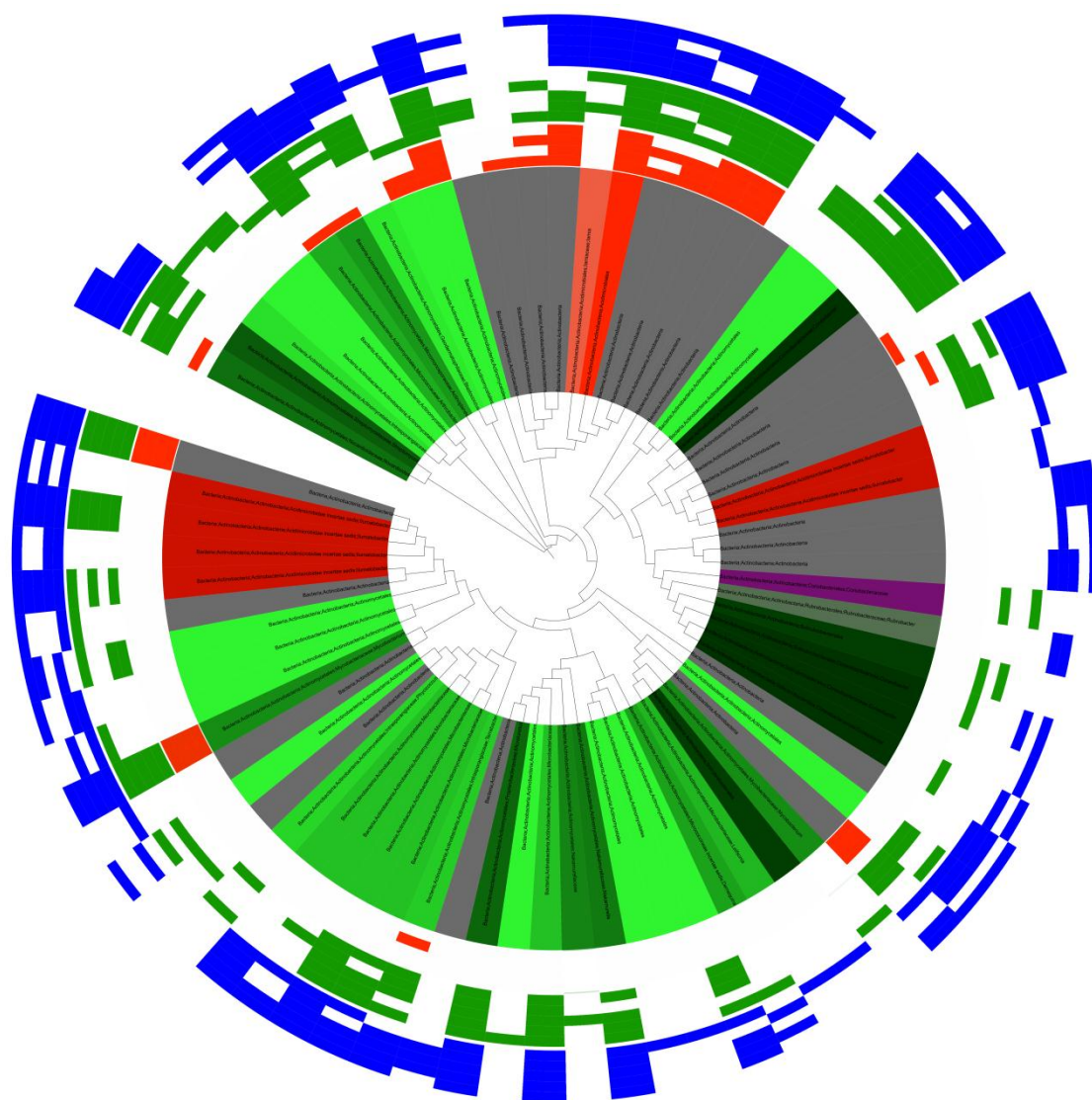
Appendix M: i) OTUs identified as group 1 Acidobacteria were predominantly found in all low and medium pH samples, however were rarely seen in the high pH samples. Two OTUs identified as group 6 Acidobacteria showed different patterns in relation to pH depending upon the OTU in question. One (top right of the tree) was seen in all samples, however the other (bottom left of the tree) was only found in a singular high pH sample. Group 2 Acidobacteria were seen in most low and medium pH samples as were group 13 and the Holophagaceae. Blue heat maps at the terminal nodes represented the high pH samples, green, the medium pH samples and red, the low pH samples.

Phylogenetic Tree Showing the Presence or Absence of Alphaproteobacterial OTUs Across all Samples (V6 – V9 16S rRNA gene hypervariable regions)



Appendix M: ii) Alphaproteobacterial lineages were predominantly found within the medium and high pH samples with a degree of endemism relating to the Rhizobiales (in contradiction to the v1 – v3 sequences). No OTUs were endemic to the low pH samples although the majority of those found in the low pH samples had limited taxonomic inference, generally to the phylum level. Radial heat maps show the presence or absence of specific OTUs, blue represented the 5 high pH samples, green the medium pH samples and red the low pH samples.

Phylogenetic Tree Showing the Presence or Absence of Actinobacterial OTUs Across all Samples (V6 – V9 16S rRNA gene hypervariable regions)

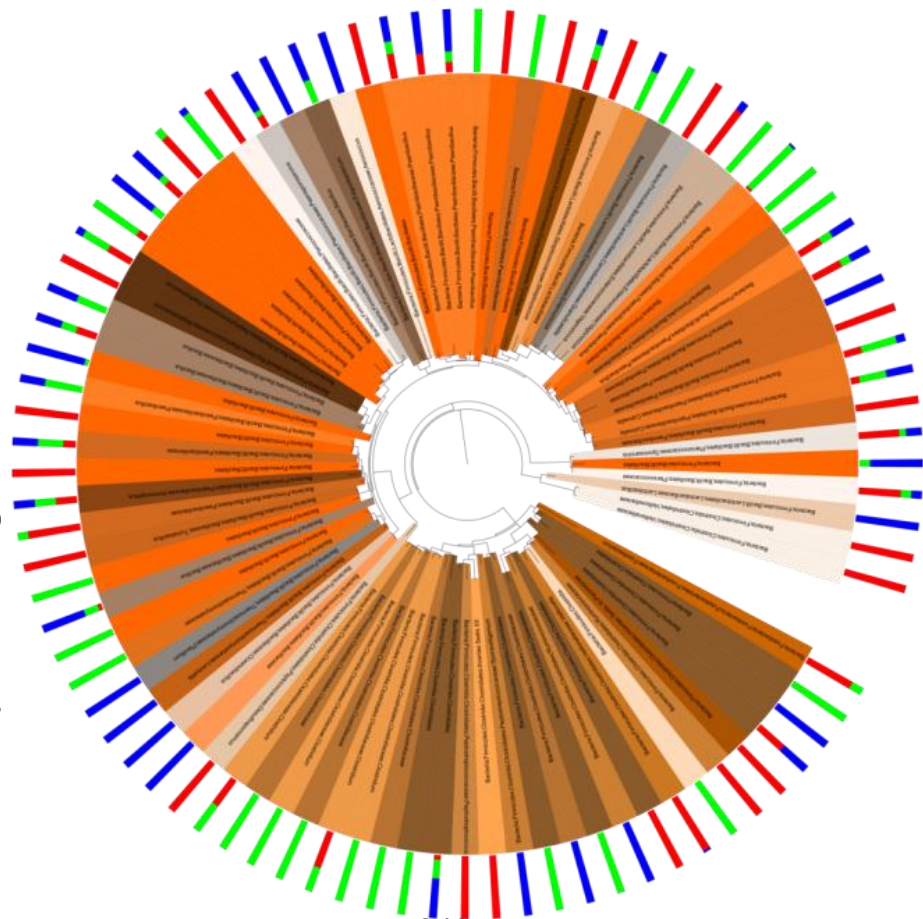


Appendix M: iii) As with the v1 – v3 shared OTU trees, many of the Actinobacterial OTUs were endemic to the medium and high pH soils, excluding those of a low pH nature. Those found within the low pH samples clustered with the Acidimicrobiales, however many of these were without detailed taxonomic information. Several OTUs identified as Acidimicrobiales were also only detected within the high pH samples thus contradicting patterns found within the V1 – V3 trees examining shared OTUs across the pH groupings. Red heat maps represent the low pH samples, green, the medium pH samples and blue, the high pH samples.

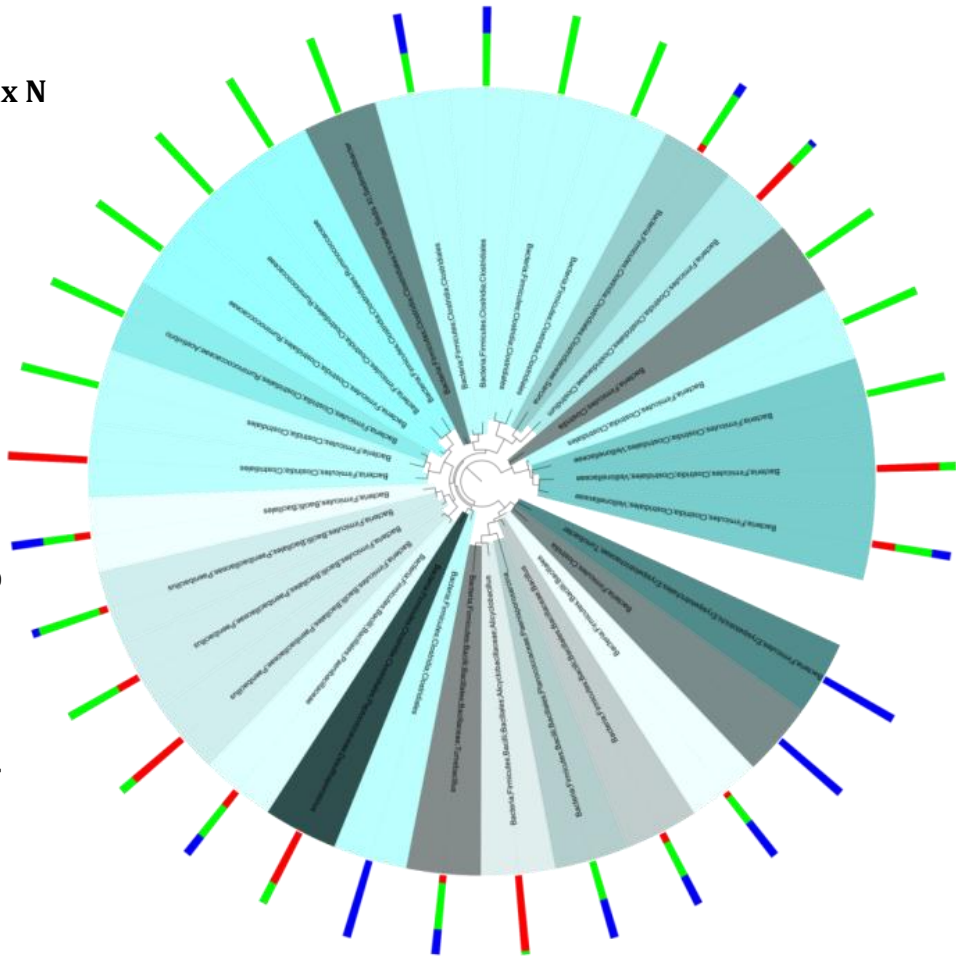
Phylogenetic placement of OTUs assigned as Firmicutes and their mean proportional abundance in low, medium and high pH soils.

Appendix N

i) V1 – V3 Regions of the 16S rRNA Gene



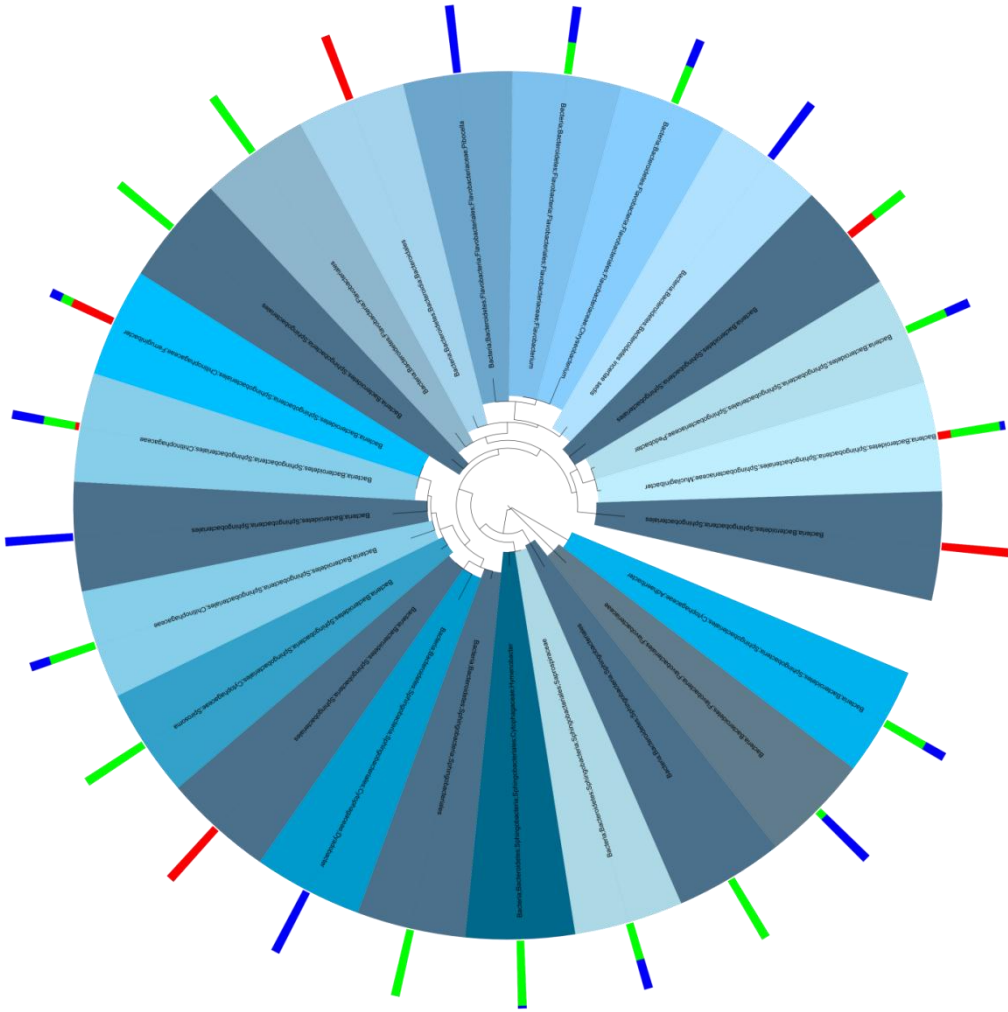
ii) V6 – V9 Regions of the 16S rRNA Gene



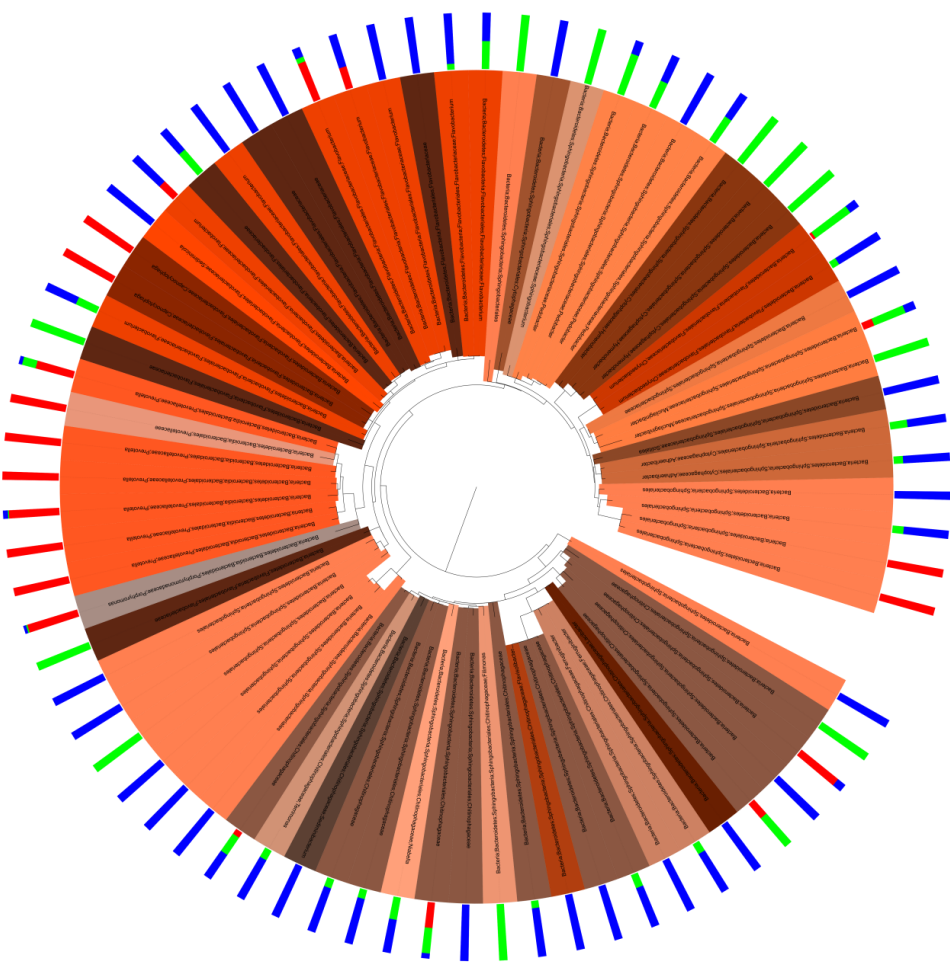
Appendix N: Phylogenetic trees showing placements of OTUs designated as Firmicutes by the RDP classifier. i) Tree constructed using V1 – V3 regions of the 16S rRNA gene at 90 % cluster similarity and ii) using the V6 – V9 regions of the 16S rRNA gene at 90 % OUT similarity. Terminal nodes show RDP classifications as close to the genus level as possible (confidence minimum = 0.8) and bar charts show the mean proportional abundance of each OTU in the low (red) medium (green) and high (blue) pH soils. Fewer OTUs were designated as Firmicutes using the V6 – V9 regions, but in both cases numerous OTUs were unique to the medium pH classes, predominantly comprising of the Clostridiales.

Phylogenetic placement of OTUs assigned as Bacteroidetes and their mean proportional abundance in low, medium and high pH soils.

iv) V6 – V9 Regions of the 16S rRNA Gene

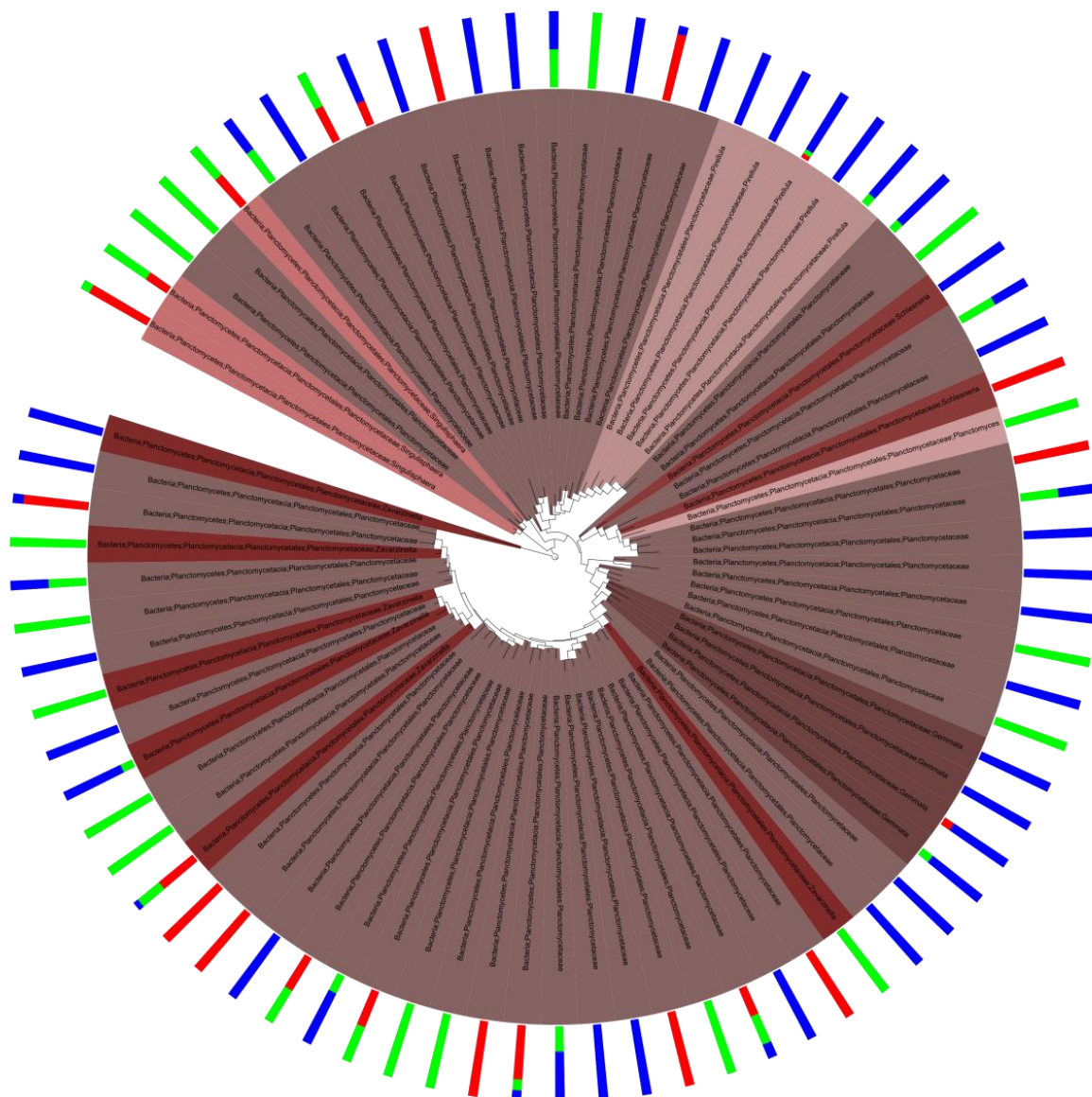


iii) V1 – V3 Regions of the 16S rRNA Gene



Appendix N: Phylogenetic placement of OTUs with taxa assignments of Bacteroidetes using the iii) V1 – V3 region sequences and iv) V6 – V9 region sequences. Terminal nodes show taxa assignments as close to the genus level as possible (RDP classifier minimum confidence = 0.8) and bar charts show the mean proportional abundance of that OTU in low (red) medium (green) and high (blue) pH soil samples. Within the V1 – V3 region tree, several OTUs were unique to the medium and low pH samples (Hymenobacter and Prevotella respectively). This finding could not be compared with the V6 – V9 tree as taxonomic assignments were not confidently assigned below the order-family level. Many of these unique OTUs were singletons and perhaps do not accurately represent the proportional abundance of these taxa using this comparably moderate level of sequencing.

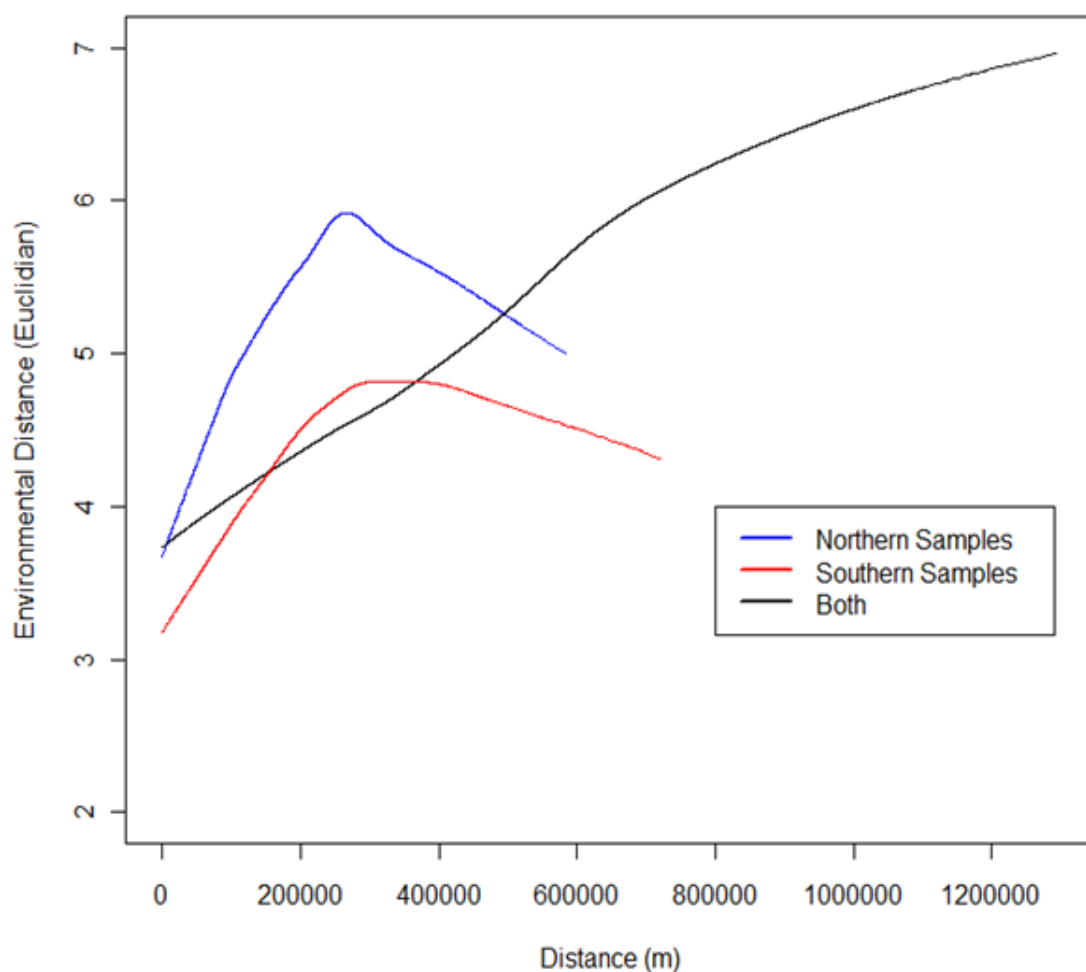
Phylogenetic Placement of OTUs Assigned as Planctomycetes and their Mean Proportional Abundance in Low, Medium and High pH Soils using the V1 – V3 Sequences.



Appendix N: v) Phylogenetic placement of OTUs classified as Planctomycetes using the RDP naïve Bayesian classifier. Principally, no Planctomycetes OTUs were detected using the V6 – V9 16S rRNA gene hypervariable regions tree, however using the V1 – V3 16S rRNA gene hypervariable region robust taxa assignments could be given to most OTUs to the family/genus level (RDP minimum confidence = 0.8). Numerous OTUs were only found in particular pH groups; however a large proportion of OTUs identified were singletons.

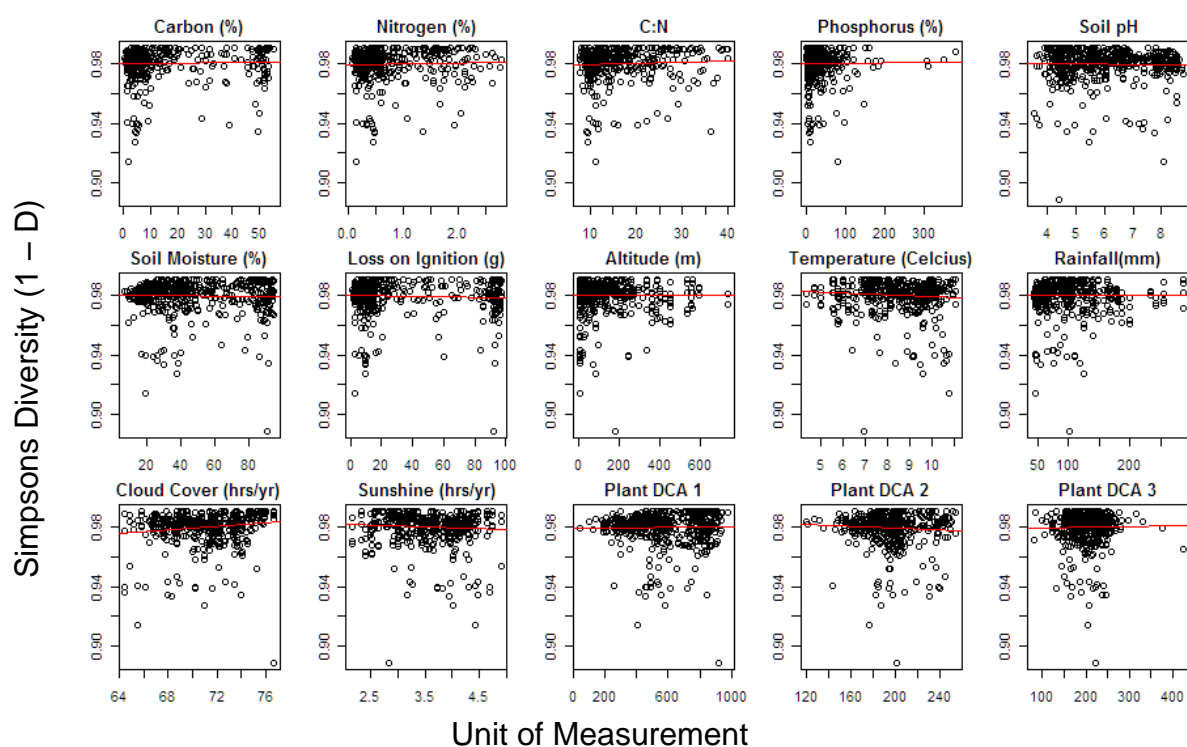
Appendix O

Environmental Dissimilarity for Northern and Southern Subsamples

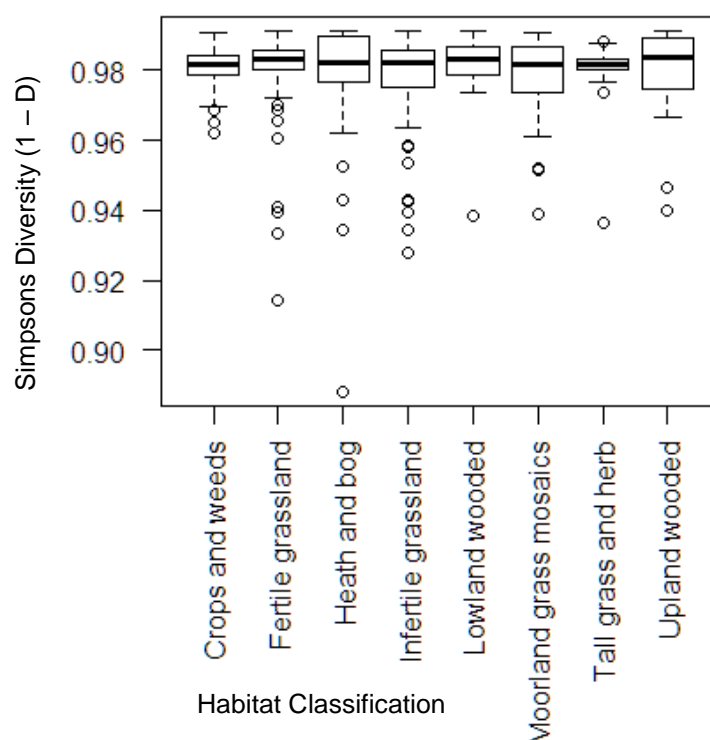


Appendix O: Environmental Dissimilarity increases with distance up to a point for the subset of Countryside Survey soil samples used for fungal t-RFLP analysis. Northern samples have a consistently higher level of environmental variability than the southern samples, however when combined, the level of environmental variability is greater than individually.

Appendix P

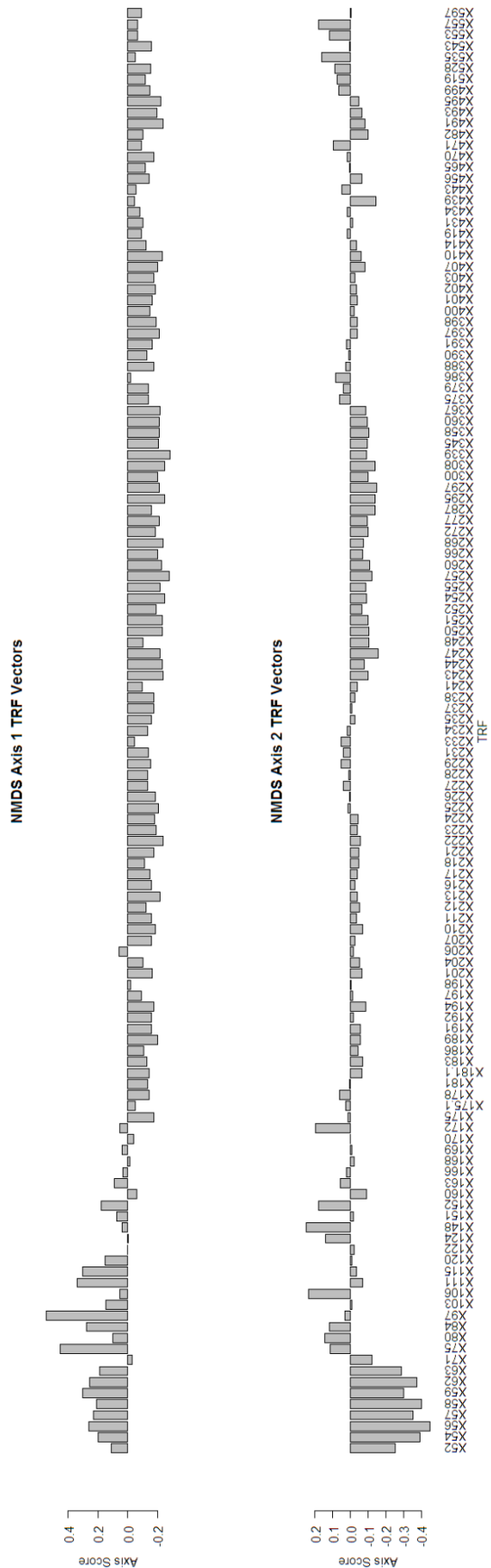


Boxplot of Fungal Diversity and Habitat Classification



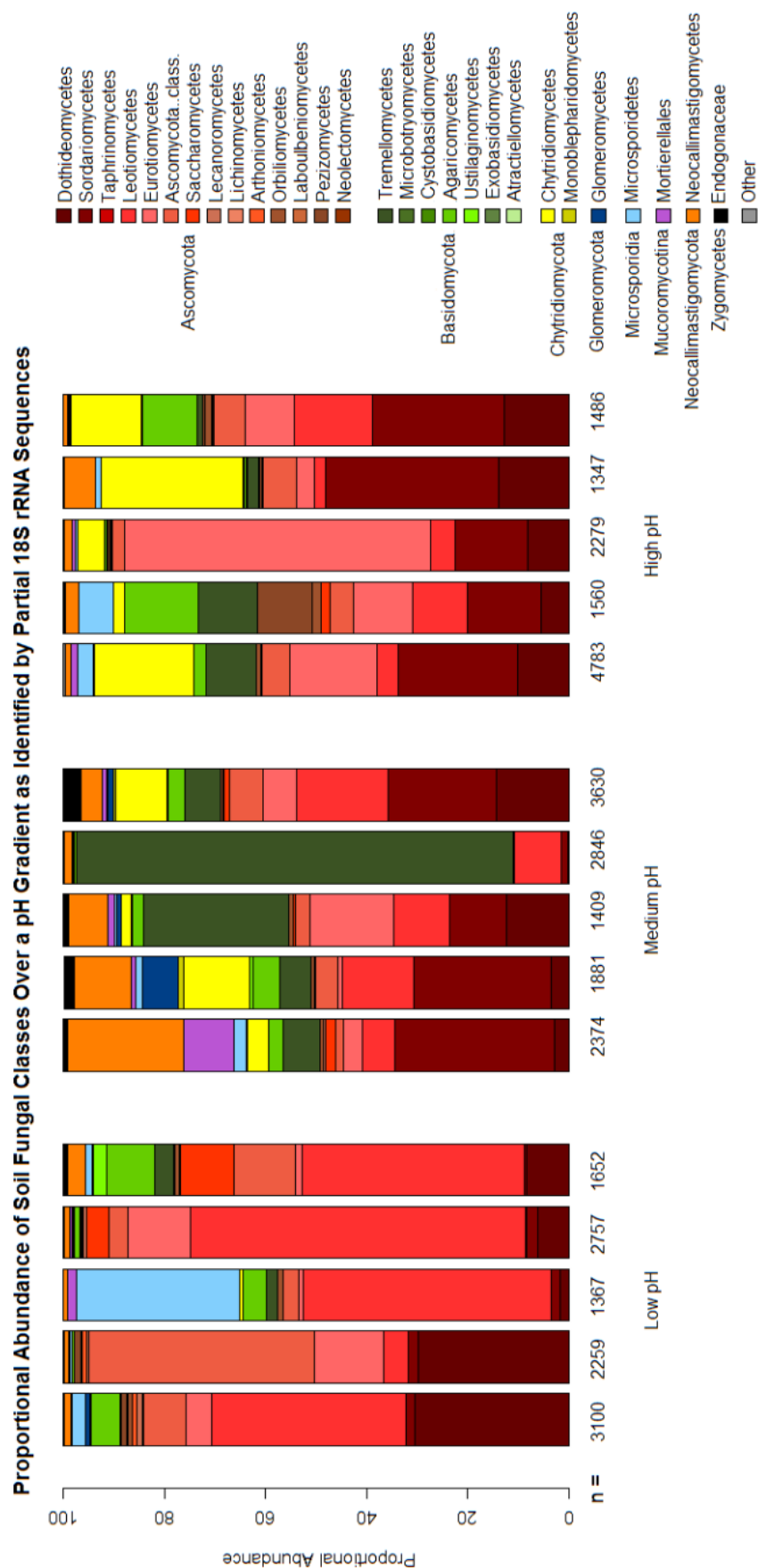
Appendix P: No Significant correlation was detected between any numerical environmental variable and fungal Diversity ($P > 0.1$). Also no significant difference was detected in terms of fungal diversity between habitat classifications. However, all calculated diversities were very high, with the majority of the data points siting between 0.96 and 1.

Appendix Q



Appendix Q: First and second axis scores for fungal TRFs across the NMDS ordination in Figure 32.

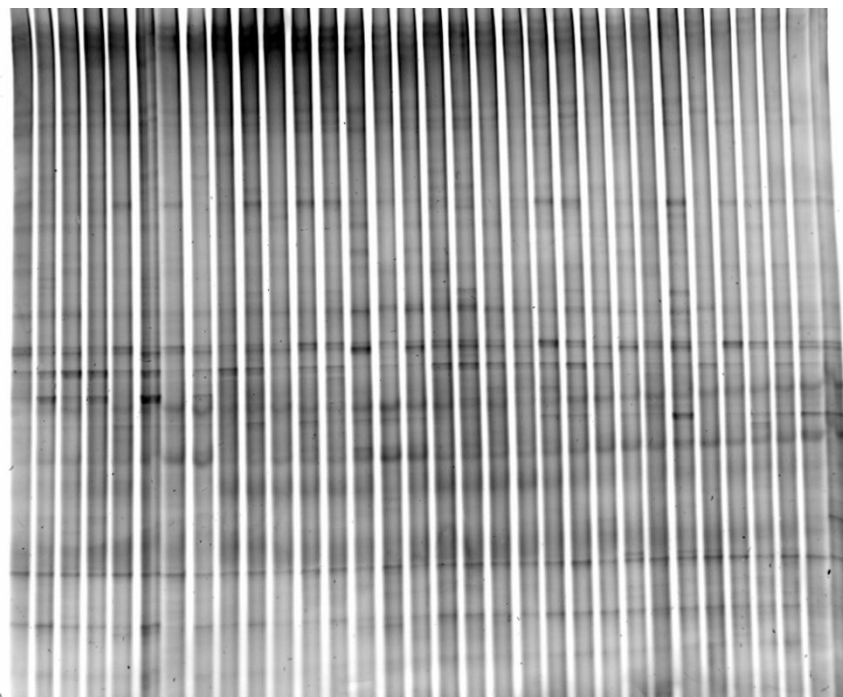
Appendix R



Appendix R: The per sample proportional abundance of fungal phyla/orders across the pH gradient as a result of 18S rRNA targeted pyrosequencing. The number of fungal sequences returned per sample are displayed under each bar and sample names increase from left to right, i.e. bar 1 = low.1, bar 2 = low.2 and so forth.

Appendix S

A)



B)



Appendix S: A) DGGE Gel of 32 *Lolium* sp. dominated soils from neutral mineral grasslands across the United Kingdom. Lane allocation is randomised to prevent spatial artefacts due to micro variations in gel formation or lateral temperature fluctuations. Bacterial community composition is comparably constant across all samples with many bands present across all lanes. B) DGGE Gel of 32 *Calluna* sp. dominated soils from acidic, organic bogs across the United Kingdom. Bacterial community composition is very variable between samples with few bands seen across all lanes. This results in a significantly greater mean dissimilarity when compared with the *Lolium* sp. dominated samples shown above (t-test $p < 0.001$).

References

- Acosta-Martinez, V., Dowd, S., Sun, Y. & Allen, V., 2008. Tag-Encoded Pyrosequencing analysis of Bacterial Diversity in a Single Soil Type Affected by Management and Land Use. *Soil Biology and Biochemistry*, 40(11): 2762-2770.
- Alder, G. H. & Dudley, R., 1994. Butterfly Biogeography and Endemism to Tropical Pacific Islands. *Biological Journal of the Linnean Society*, 52(2): 151 - 162.
- Allen, E. et al. 1995. Patterns and Regulation of Mycorrhizal Plant and Fungal Diversity. *Plant Soil*, 170: 47 - 62.
- Allison, S. D., Hanson, C. A. & Treseder, K. K., 2007. Nitrogen Fertilization Reduces Diversity and Alters Community Structure of Active Fungi in Boreal Ecosystems. *Soil Biology and Biochemistry*, 39(8): 1878-1887.
- Allison, S. M. & Prosser, J., 1993. Ammonia Oxidation at Low pH by Attached Populations of Nitrifying Bacteria. *Soil Biology and Biochemistry*, 25(7): 935-941.
- Altschul, S. F. et al. 1990. Basic Local Alignment Search Tool. *Journal of Molecular Biology*, 215(3): 403 - 410.
- Anderson, I., Campbell, C. & Prosser, J., 2003. Potential Bias of Fungal 18S rDNA and Internal Transcribed Spacer Polymerase Chain Reaction Primers for Estimating Fungal Biodiversity in Soil. *Environmental Microbiology*, 5(1): 36 - 47.
- Anderson, I. C. & Cairney, J. W. G., 2004. Diversity and Ecology of Soil Fungal Communities: Increased Understanding Through the Application of Molecular Techniques. *Environmental Microbiology*, 6(8): 769-779.
- Andren, O. & Balandreau, J., 1999. Biodiversity and Soil Functioning - From Black Box to Can of Worms? *Applied Soil Ecology*, 13: 100-108.
- Arnell, N., Brown, R. P. & Reynard, N. S., 1990. Impact of Climatic Variability and Change on River Flow Regimes in the UK, Wallingford: Institute of Hydrology.
- Avis, P. G., Dickie, I. A. & Mueller, G. M., 2006. A "Dirty" Business: Testing the Limitations of Terminal Restriction Fragment Length Polymorphism (TRFLP) Analysis of Soil Fungi. *Molecular Ecology*, 15(3): 873-882.
- Azaizeh, H. A., Marschner, H. & Romheld, V., 1995. Effects of Vesicular-Arbuscular Mycorrhizal Fungus and Other Soil Microorganisms on Growth, Mineral Nutrient Acquisition and Root Exudation of Soil-Grown Maize Plants. *Mycorrhiza*, 5: 321-327.

- Baath, E. & Anderson, T. H., 2003. Comparison of Soil Fungal/Bacterial Ratios in a pH Gradient using Physiological and PLFA-Based Techniques. *Soil Biology and Biochemistry*, 35: 955 - 963.
- Bais, H. P. et al. 2006. The Role of Root Exudates in Rhizosphere Interactions with Plants and Other Organisms. *Annual Review of Plant Biology*, 57: 233-266.
- Bakkenes, M. et al. 2002. Assessing the Effects of Forecasted Climate Change on the Diversity and Distribution of European Higher Plants for 2050. *Global Change Biology*, 8(4): 390-407.
- Bakri, Y., Masson, M. & Thornart, P., 2010. Isolation and Identification of Two New Fungal Strains for Xylanase Production. *Applied Biochemistry and Biotechnology*, 162(6): 1626-1634.
- Bale, J. et al. 1997. Life Strategies of Arctic Terrestrial Arthropods. In: S. M. M. Woodin, ed. *Ecology of Arctic Environments*. Cambridge: Blackwell Science: 137 - 164.
- Ball, D. F. & Williams, W. M., 1968. Variability of Soil Chemical Properties in Two Uncultivated Brown Earths. *European Journal of Soil Science*, 19(2): 379 - 391.
- Banwart, S., 2011. Save our Soils. *Nature*, 474: 151-152.
- Barber, S. A., 1995. Soil Nutrient Bioavailability: A Mechanistic Approach. 2nd ed. Wisconsin: Wiley & Sons.
- Bardgett, R. D., 2005. The Biology of Soil: A Community and Ecosystem Approach. 1st ed. Oxford: Oxford University Press.
- Bardgett, R. D., Usher, M. B. & Hopkins, D. W., 2005. *Biological Diversity and Function in Soils*. 1st ed. Cambridge: Cambridge University Press.
- Bardgett, R. D. & McAlister, E., 1999. The Measurement of Soil Fungal:Bacterial Biomass ratios as an Indicator of Ecosystem Self-Regulation in Temperate Meadow Grasslands. *Biology and Fertility of Soils*, 29: 282-290.
- Baudoin, E., Benizri, E. & Guckert, A., 2003. Impact of Artificial Root Exudates on the Bacterial Community Structure in Bulk Soil and Maize Rhizosphere. *Soil Biology and Biochemistry*, 35(9): 1183-1192.
- Baxter, J. W. & Dighton, J., 2005. Diversity-Functioning Relationships in Ectomycorrhizal Fungal Communities. In: P. Oudemans, ed. *The Fungal Community: Its Organisation and Role in the Ecosystem*. Raton: CRC Press: 383-399.
- Beales, N., 2004. Adaptation of Microorganisms to Cold Temperatures, Weak Acid Preservatives, Low pH, and Osmotic Stress: A Review. *Comprehensive Reviews in Food Science and Food Safety*, 3(1): 1-20.

- Beare, M. H. et al. 1995. A Hierarchical Approach to Evaluating the Significance of Soil Biodiversity to Biogeochemical Cycling. *Plant and Soil*, 170(1): 5-22.
- Bennison, G. M. & Wright, A. E., 1969. *The Geological History of the British Isles*. London: Edward Arnold.
- Bent, S. J., Pierson, J. D. & Forney, L. J., 2007. Measuring Species Richness Based on Microbial Community Fingerprints: The Emperor Has No Clothes. *Applied and Environmental Microbiology*, 73(7): 2399-2401.
- Berg, B. & McClaugherty, C., 2008. *Plant Litter, Decomposition, Humus Formation, Carbon Sequestration*. 2nd ed. Berlin: Springer-Verlag.
- Bergerow, D., Nilsson, H., Unterseher, M. & Maier, W., 2010. Current State and Perspectives of Fungal DNA Barcoding and Rapid Identification Procedures. *Applied Microbiology and Biotechnology*, 87(1): 99-108.
- Boer, W., Folman, L. B., Summerbell, R. C. & Boddy, L., 2005. Living in a Fungal World: Impact of Fungi on Soil Bacterial Niche Development. *FEMS Microbiology Reviews*, 29(4): 795-811.
- Boer, W. et al. 1991. Nitrification at Low pH by Aggregated Chemolithotrophic Bacteria. *Applied and Environmental Microbiology*, 57(12): 3600-3604.
- Boer, W. & Kowalchuk, G. A., 2001. Nitrification in Acid Soils: Microorganisms and Mechanisms. *Soil Biology and Biochemistry*, 33(7): 853-866.
- Bohn, H. L., 1976. Estimate of Organic Carbon in World Soils. *Soil Science Society of America Journal*, 40(3): 468-470.
- Bohn, H., McNeal, B. & O'Connor, G., 2001. *Soil Chemistry*. III ed. New York: John Wiley & Sons, Inc..
- Bolin, B., 1970. The Carbon Cycle. *Scientific American*, 223(3).
- Booth, M. S., Stark, J. M. & Rastetter, E., 2005. Controls on Nitrogen Cycling in Terrestrial Ecosystems: A Synthetic Analysis of Literature Data. *Ecological Monographs*, 75(2): 139-157.
- Borcard, D., Legendre, P., Drapeau, P., 1992. Partialling out the Spatial Component of Ecological Variation. *Ecology*, 73(3): 1045-1055.
- Bowers, R. M., Sullivan, A. P., Costello, E. K., Knight, R., Fierer, N., 2011. Sources of Bacteria in Outdoor Air Across Cities in the Midwestern United States. *Applied and Environmental Microbiology*, Online First
- Brady, N. C. & Weil, R. R., 1996. *The Nature and Properties of Soils*. 2nd ed. New Jersey: Prentice-Hall Inc.

- Braid, M., Daniels, L. M. & Kitts, C. L., 2003. Removal of PCR Inhibitors From soil DNA by Chemical Flocculation. *Journal of Microbiological Methods*, 52(3): 389-393.
- Brodie, E., Edwards, S. & Clipson, N., 2003. Soil Fungal Community Structure in a Temperate Upland Grassland Soil. *FEMS Microbial Ecology*, 45(2): 105-114.
- Broekling, C. et al. 2007. Root Exudates Regulate Soil Fungal Community Composition and Diversity. *Applied and Environmental Microbiology*, 74(3): 738 - 744.
- Brookes, P. C., Powlson, D. S. & Jenkinson, D. S., 1984. Phosphorus in the Soil Microbial Biomass. *Soil Biology and Biochemistry*, 16(2): 169-175.
- Buee, M. et al. 2009. 454 Pyrosequencing Analyses of Forest Soils Reveal an Unexpectedly High Fungal Diversity. *New Phytologist*, 184: 449 - 456.
- Buol, S. W., Southard, R. J., Graham, R. C. & McDaniel, P. A., 2003. *Soil Genesis and Classification*. 5th ed. Ames: Blackwell Publishing.
- Cai, P. et al. 2007. Amplification of Plasmid DNA Bound on Soil Colloidal Particles and Clay Minerals by Polymerase Chain Reaction. *Journal of Environmental Sciences*, 19(11): 1326:1329.
- Caporaso, J. G. et al. 2010. PyNAST. *Bioinformatics*, 26(2): 266 - 267.
- Caporaso, J. G. et al. 2010. QIIME Allows Analysis of High-Throughput Community Sequencing Data. *Nature Methods*, 7: 335 - 336.
- Carey, P. D. et al. 2008. Countryside Survey: UK Headline Messages from 2007, : NERC/Centre for Ecology and Hydrology.
- Carlile, W. R., 1995. *Control of Crop Diseases*. 2nd ed. Cambridge: Cambridge University Press.
- Caron, D. A. et al. 2009. Protists are Microbes Too: a Perspective. *The ISME Journal*, 3: 4 - 12.
- Case, R. J. et al. 2007. Use of 16S rRNA and rpoB Genes as Molecular Markers for Microbial Ecology Studies. *Applied and Environmental Microbiology*, 73(1): 278-288.
- Chapela, I. H., Rehner, S. A., Schultz, T. R. & Mueller, U. G., 1994. Evolutionary History of the Symbiosis Between Fungus-Growing Ants and Their Fungi. *Science*, 266(5191): 1691-1694.
- Chao, A., 1984. Non-Parametric Estimation of the Number of Classes in a Population. *Scandinavian Journal of Statistics*. 11: 165 - 270

Chistoserdova, L., 2010. Recent Progress and new Challenges in Metagenomics for Biotechnology. *Biotechnology Letters*, 32: 1351-1359.

Cho, J. C. & Tiedje, J. M., 2000. Biogeography and Degree of Endemicity in Fluorescent *Pseudomonas* Strains in Soil. *Applied and Environmental Microbiology*, 66(12): 5448-5456.

Christensen, M., 1989. A View of Fungal Ecology. *Mycologia*, 81: 1-19.

Chu, H. et al. 2010. Soil Bacterial Diversity in the Arctic is not Fundamentally Different from that found in other Biomes. *Environmental Microbiology*, 12(11): 2998 - 3006.

Chu, H., Neufeld, J., Walter, V. & Grogan, P., 2011. The Influence of Vegetation Type on the Dominant Soil Bacteria, Archaea, and Fungi in Low Arctic Tundra Landscape. *Soil Science Society of America Journal*, 75(5): 1756 - 1765.

Coates, J. D., Ellis, D., Gaw, C. & Lovley, D. R., 1999. *Geothrix fermentans* gen. nov., sp. nov., a novel Fe(III)-Reducing Bacterium from a Hydrocarbon-Contaminated Aquifer. *International Journal of Systematic Bacteriology*, 49: 1615-1622.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed) Lawrence Erlbaum Associates, New Jersey, USA.

Cole, L., Bardgett, R. D., Ineson, P. & Adamson, J. K., 2002. Relationships Between Enchytraeid Worms (*Oligochaeta*), Climate Change, and the Release of Dissolved Organic Carbon from Blanket Peat in Northern England. *Soil Biology and Biochemistry*, 35(5): 599-607.

Collins, F., Morgan, M. & Patrinos, A., 2003. The Human Genome Project: Lessons from Large-Scale Biology. *Science*, 300: 286-290.

Comas, L. H., Tuininga, A. R. & Callahan, H. S., 2010. Advancing our Current Understanding of Plant-Fungal Symbioses: Bridging Scales from Local to Global. *New Phytologist*, 185: 871-873.

International Soil Metagenome Sequencing Consortium, 2012. Terragenome. [Online]

Available at: <http://www.terrigenome.org/about/>

[Accessed 11 1 2010].

Cornelissen, J. H. et al. 1999. Leaf Structure and Defence Control Litter Decomposition Rate Across Species and Life Forms in Regional Floras on two Continents. *New Phytologist*, 143(1): 191-200.

- Costa, R. et al. 2006. Effects of Site and Plant Species on Rhizosphere Community Structure as Revealed by Molecular Analysis of Microbial Guilds. *FEMS Microbiology Ecology*, 56: 236-249.
- Cotter, P. D. & Hill, C., 2003. Surviving the Acid Test: Responses of Gram-Positive Bacteria to Low pH. *Microbiology and Molecular Biology Reviews*, 67(3): 429-453.
- Cox, B. & Moore, P., 2010. *Biogeography: An Ecological and Evolutionary Approach*. 8th ed. Danvers: Wiley & Sons.
- Crocker, R. I. & Major, J., 1955. Soil Development in Relation to Vegetation and Surface Age at Glacier Bay, Alaska. *Journal of Ecology*, 43(2): 427-448.
- Cullings, K. & Hanely, J., 2010. Dwarf Mistletoe Effects on Soil Basidiomycete Community Structure, Soil Fungal Functional Diversity, and Soil Enzyme Function: Implications for Climate Change. *Soil Biology and Biochemistry*, 42(11): 1976 - 1981.
- Curtis, T. P., William, S. T. & Scannell, J. W., 2002. Estimating Prokaryotic Diversity and its Limits. *PNAS*, 99(16): 10494-10499.
- Dalal, R. C., 2001. Acidic Soil pH, Aluminium and Iron Affect Organic Carbon Turnover in Soils. *NEE Workshop Proceedings*. 111-114.
- Daniell, T. J., Husband, R., Fitter, J. P. & Young, W., 2006. Molecular Diversity of Arbuscular Mycorrhizal Fungi Colonising Arable Crops. *FEMS Microbiology Ecology*, 36(2): 203-209.
- Darwin, D., 1859. *On the Origin of Species by Means of Natural Selection*. Plymouth: Broadview Press.
- Davey, H. M. & Kell, D. B., 1996. Flow Cytometry and Cell Sorting of Heterogeneous Microbial Populations: The Importance of Single-Cell Analysis. *Microbiology and Molecular Biology Reviews*, 60(4): 641-696.
- Davey, E, M. & O'Toole, G. A., 2000. Microbial Biofilms: from Ecology to Molecular Genetics. *Microbiology and Molecular Biology Reviews*, 64(4): 847-867.
- Davidson, E. A. & Janssens, I. A., 2006. Temperature Sensitivity of Soil Carbon Decomposition and Feedbacks to Climate Change. *Nature*, 440: 165-173.
- De Bono, M., 2003. Molecular Approaches to Aggregation Behaviour and Social Attachment. *Journal of Neurobiology*, 54(1): 78-92.
- De Vos, C. R., Lubberding, H. J. & Bienfait, H. F., 1986. Rhizosphere Acidification as a Response to Iron Deficiency in Bean Plants. *Plant Physiology*, 81:842-846.

- De'ath, G., 2002. Multivariate Regression Trees: A New Technique for Modelling Species-Environment Relationships. *Ecological Society of America*, 83(4): 1105-1117.
- Decaens, T. et al. 2006. The Values of Soil Animals for Conservation Biology. *European Journal of Soil Biology*, 42: 23 - 38.
- d'Enfert, C., 2009. Hidden Killers: Persistence of Opportunistic Fungal Pathogens in the Human Host. *Current Opinion in Microbiology*, 12(4): 358-364.
- Derakshani, M., Lukow, T., Liesack, W., 2001, Novel Bacterial Lineages at the (Sub)Division Level as Detected by Signature Nucleotide-Targeted Recovery of 16S rRNA Genes from Bulk Soil and Rice Roots of Flooded Rice Microcosms. *Applied and Environmental Microbiology*, 67(2):623 - 631
- Dequiedt, S. et al. 2011. Biogeographical Patterns of Soil Molecular Microbial Biomass as Influenced by Soil Characteristics and Management. *Global Ecology and Biogeography*, 20: 641-652.
- DeSantis, T. Z. et al. 2006. Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Applied and Environmental Microbiology*, 72: 5069 - 5072.
- Diaz, S., Marcelo, C. & Fernando, C., 1998. Plant Functional Traits and Environmental Filters at a Regional Scale. *Journal of Vegetation Science*, 9: 113-122.
- Dinkelaker, B., Romheld, V. & Marschner, H., 2006. Citric Acid Excretion and Precipitation of Calcium Citrate in the Rhizosphere of White Lupin (*Lupinus albus* L.). *Plant, Cell and Environment*, 12(3): 285-292.
- Dirnbock, T. & Dullinger, S., 2004. Habitat Distribution Models, Spatial Autocorrelation, Functional Traits and Dispersal Capacity of Alpine Plant Species. *Journal of Vegetation Science*, 15: 77-84.
- Dolan, J. R., 2006. Microbial Biogeography?. *Journal of Biogeography*, 33: 199-200.
- Doolittle, R. F. et al. 1996. Determining Divergence Times of the Major Kingdoms of Living Organisms with a Protein Clock. *Science*, 271(5248): 470-477.
- Dowd, S. E. & Callaway, T. R., 2008. Evaluation of the Bacterial Diversity in the Faeces of Cattle using 16S rDNA Bacterial tag-encoded FLX Amplicon Pyrosequencing (bTEFAP). *BMC Microbiology*, 8: 125.
- Dray, S., Legendre, P. & Peres-Neto, P., 2006. Spatial Modelling: A Comprehensive Framework for Principal Coordinate Analysis of Neighbour Matrices (PCNM). *Ecological Modelling*, 196(2-3): 483-493.

- Dufrene, M., & Legendre, P (1996). Species Assemblages and Indicator Species: The Need for a Flexible and Asymmetrical Approach. *Ecological Monographs*, 67:345-366
- Dunbar, J. et al. 1999. Levels of Bacterial Community Diversity in Four Arid Soils Compared by Cultivation and 16S rRNA Gene Cloning. *Applied and Environmental Microbiology*, 65: 1662 - 1669.
- Edgar, R., 2004. MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput. *Nucleic Acids Research*, 32(5): 1792 - 97.
- Edgar, R. C., 2010. Search and Clustering Orders of Magnitude Faster than BLAST. *Bioinformatics*, 26(19): 2460 - 2461
- Emmett, B. A. et al. 2007. Countryside Survey: Soils Report from 2007, s.l.: s.n.
- Ettema, C. H. & Wardle, D. A., 2002. Spatial Soil Ecology. *Trends in Ecology and Evolution*, 17(4): 177-188.
- Feinstein, L. M., Sul, W. J. & Blackwood, C. B., 2009. Assessment of Bias Associated with Incomplete Extraction of Microbial DNA from Soil. *Applied and Environmental Microbiology*, 75(16): 5428 - 5433.
- Felske, A. et al. 1997. Ribosome analysis Reveals Prominent Activity of an Uncultured Member of the class Actinobacteria in grassland soils. *Microbiology*, 143(9): 2983-2989.
- Ferrari, B., Binnerup, S. J. & Gillings, M., 2005. Micro-colony Cultivation on a Soil Substrate Membrane System Selects for Previously Uncultured Soil Bacteria. *Applied and Environmental Microbiology*, 71(12): 8714-8720.
- Fierer, N. et al. 2007. Metagenomic and Small-Subunit rRNA Analysis Reveal the Genetic Diversity of Bacteria, Archaea, Fungi and Viruses in Soil. *Applied and Environmental Microbiology*, 73(21): 7059:7066.
- Fierer, N., 2007. Tilting at Windmills: a Response to a Recent Critique of Terminal Restriction Fragment Length Polymorphism Data. *Applied and Environmental Microbiology*, 73(24): 8041:8042.
- Fierer, N., 2008. Microbial Biogeography: Patterns in Microbial Diversity Across Space and Time. In: K.Zengler, ed. *Accessing uncultivated microorganisms: from the environment to organisms and genomes and back*. Washington: ASM Press: 95 - 115.
- Fierer, N., Bradford, M. & Jackson, R., 2007. Towards an Ecological Classification of Soil Bacteria. *Ecology*, 88(6): 1354-1364.
- Fierer, N. et al. 2007. Metagenomic and Small-Subunit rRNA Analysis Reveal the Genetic Diversity of Bacteria Archaea Fungi and Viruses in Soil. *Applied and Environmental Microbiology*, 72(21): 7059-7066.

- Fierer, N. & Jackson, R. B., 2006. The Diversity and Biogeography of Soil Bacterial Communities. *PNAS*, 103(3): 626-631.
- Fierer, N., Schimel, J. & Holden, P. A., 2003. Variations in Microbial Community Composition Through two Soil Depth Profiles. *Soil Biology and Biochemistry*, 35(1): 167-176.
- Fierer, N. & Schimel, J. P., 2002. Effects of Drying and Rewetting Frequency on Soil Carbon and Nitrogen Transformations. *Soil Biology and Biochemistry*, 34(6): 777-787.
- Fierer, N. et al. 2009. Global Patterns in Below ground Communities. *Ecology Letters*, 12(11): 1238-1249.
- Finlay, J. B., 2002. Global Dispersal of Free-Living Microbial Eukaryote Species. *Science*, 296(5570): 1061-1063.
- Fox, G. E., Wisotzkey, J. D. & Jurtshuk, P., 1992. How close is close: 16S rRNA Sequence Identity May Not be Sufficient to Guarantee Species Identity. *IJSEM*, 42(1): 166-170.
- Franklin, R. B. & Mills, A. L., 2003. Multi-Scale Variation in Spatial Heterogeneity for Microbial Community Structure in an Eastern Virginia Agricultural Field. *FEMS Microbiology Ecology*, 44: 335 - 346.
- Franzluebbers, A. J., Langdale, G. W. & Schomberg, H. H., 1999. Soil Carbon, Nitrogen and Aggregation in Response to Type and Frequency of Tillage. *Soil Science Society of America*, 63(2): 349-355.
- Freeman, C., Ostle, N. & Kang, H., 2001. An Enzymatic 'Latch' on a Global Carbon Store. *Nature*, 409: 149.
- Fulthorpe, R. R., Roesch, F. W., Riva, A. & Tripplet, E. W., 2008. Distantly Sampled Soils Carry Few Species in Common. *The ISME Journal*, 2: 901-910.
- Gans, J., Wolinsky, M. & Dunbar, J., 2005. Computational Improvements Reveal Great Bacterial Diversity and High Metal Toxicity in Soil. *Science*, 309(5739): 1387-1390.
- Gardes, M. & Dahlberg, A., 1996. Mycorrhizal Diversity in Arctic and Alpine Tundra: An Open Question. *New Phytologist*, 133(1): 147-157.
- Garnier-Sillam, E., Hariyanto, S. & Bourezguy, Y., 1999. Humic Substances in Peat. *Analysis*, 27(5): 405-408.
- Gianinazzi-Pearson, V., 1996. Plant Cell Responses to Arbuscular Mycorrhizal Fungi: Getting to the Roots of the Symbiosis. *The Plant Cell*, 8(10): 1871-1883.
- Giller, P. S., 1996. The Diversities of Soil Communities, the "Poor man's" tropical rainforest. *Biodiversity and Conservation*, 5: 135 - 168.

- Girvan, M. S. et al. 2003. Soil Type is the Primary Determinant of the Composition of the Total and Active Bacterial Communities in Arable Soils. *Applied and Environmental Microbiology*, 69(3): 1800-1809.
- Gomes, N. et al. 2003. Dynamics of Fungal Communities in Bulk Maize Rhizosphere Soil in the Tropics. *Applied and Environmental Microbiology*, 69(7): 3758-3766.
- Gontang, E. A., Fenical, W. & Jensen, P. R., 2007. Phylogenetic Diversity of Gram-Positive Bacteria Cultured from Marine Sediments. *Applied and Environmental Microbiology*, 73(10): 3272-3282.
- Gosling, P., Hodge, A., Goodlass, G., Bending, G. D., 2006. Arbuscular Mycorrhizal Fungi and Organic Farming. *Agriculture, Ecosystems & Environment*, 113: 17 – 35.
- Govaerts, B. et al. 2008. Long-term Consequences of Tillage, Residue Management, and Crop Rotation on Selected soil Micro-flora Groups in the Subtropical Highlands. *Applied Soil Ecology*, 23(3): 197-210.
- Green, J. & Bohannan, J. M., 2006. Spatial Scaling of Microbial Biodiversity. *Trends in Ecology and Evolution*, 21(9): 501-507.
- Griffith, D., 2003. What is Spatial Autocorrelation? In: *Spatial Autocorrelation and Spatial Filtering*. New York: Springer-Verlag: 3 - 6.
- Griffiths, B. S. et al. 2000. Ecosystem Response of Pasture Soil Communities to Fumigation-induced Microbial Diversity Reductions: an Examination of the Biodiversity-Ecosystem Functioning Relationship. *OIKOS*, 90(2): 279 - 294.
- Griffiths, R. I., Thomson, B.C., James, P.L., Bell, T., Bailey, M., Whitley, A.S., 2011. The bacterial Biogeography of UK soils. *Environmental Microbiology*, 13(6): 1642 - 1654
- Griffiths, R. I., Whiteley, A. S., O'Donnell, A. G. & Bailey, M. J., 2003. Influence of Depth and Sampling Time on Bacterial Community Structure in an Upland Grassland Soil. *FEMS Microbiology Ecology*, 43(1): 35-43.
- Griffiths, R. I., Whiteley, A. W., O'Donnell, A. & Bailey, M. J., 2000. Rapid Method for Coextraction of DNA and RNA from Natural Environments for Analysis of Ribosomal DNA- and rRNA-Based Microbial community Composition. *Applied and Environmental Microbiology*, 66(12): 5488-5491.
- Grinstead, M. J., Headley, M. J., White, R. E. & Nye, P. H., 1982. Plant-Induced Changes in the Rhizosphere of Rape (*Brassica Napus* Var. Emerald) seedlings: pH Change and the Increase in P Concentrations in the Soil Solutions. *New Phytologist*, 91: 19-29.

- Gubry-Rangin, C., Hai, B., Quince, C., Engel, M., Thomson, B.C., James, P.L., Schlöter, M., Griffiths, R.I., Prosser, J.I., Nicol, G.W., 2011. Niche Specialization of Terrestrial Archaeal Ammonia Oxidizers. *PNAS* 108(52): 21206-21211.
- Gutell, R. R., 1993. Collection of Small Subunit (16S- and 16S -like) Ribosomal RNA Structures. *Nucleic Acids Research*, 21(13): 3051 - 3054.
- Gutell, R. R., 1993. Comparative Studies of RNA: Inferring Higher-Order Structure from Patterns of Sequence Variation. *Current Opinion in Structural Biology*, 3(3): 313-322.
- Haas, B. J. et al. 2011. Chimeric 16S rRNA Sequence Formation and Detection in Sanger and 454 - Pyrosequenced PCR Amplicons. *Genome Research*, 21(3): 494 - 504.
- Hagedorn, S. A., 1994. The National Biological Impact Assessment Program and the Public Perception of Biotechnology. *Science, Technology and Society*, 14: 24-27.
- Haichar, F. Z. et al. 2008. Plant Host Habitat and Root Exudates Shape Soil Bacterial Community Structure. *The ISME Journal*, 2: 1221-1230.
- Halling, R., 2001. Ectomycorrhizae Co-evolution, Significance and Biogeography. *Annals of the Missouri Botanical Garden*, 88(1): 5-13.
- Hammel, K. E., 1997. Fungal Degradation of Lignin. In: G. Cadish & K. E. Giller, eds. *Driven by Nature: Plant Litter Quality and Decomposition*. Wallingford: CAN International, pp. 33-45.
- Hardin, G., 1960. The Competitive Exclusion Principle. *Science*, 131(3409): 1292 – 1297.
- Harris, J. A., Grogan, P. & Hobbs, R. J., 2005. Restoration Ecology and the Role of Soil Biodiversity. In: R. D. Bardgett, ed. *Biological Diversity and Function in Soils*. Cambridge: Cambridge University Press: 319-342.
- Hartman, W. H., Richardson, C. J., Vilgalys, R. & Bruland, G. L., 2008. Environmental and Anthropogenic Controls over Bacterial Communities in Wetland Soils. *PNAS*, 105: 17842 - 17847.
- Hattenschwiler, S., Tiunov, A. & Scheu, S., 2005. Biodiversity and Litter Decomposition in Terrestrial Ecosystems. *Annual Review of Ecology, Evolution and Systematics*, 36: 191 - 218.
- Hawksworth, D., 2001. The Magnitude of Fungal Diversity: the 1.5 Million Species Estimate Revisited. *Mycological Research*, 105: 1422 - 1432.
- Hawksworth, D., 2004. Fungal Diversity and its Implications for Genetic Resource Collections. *Studies in Mycology*, 50: 9 - 18.

- Haygarth, P. M. & Ritz, K., 2009. The Future of Soils and Land Use in the UK: Soil Systems for the Provision of Land-Based Ecosystem Services. *Land Use Policy*, 26(1): 187-197.
- Heckman, D. et al. 2001. Molecular Evidence for the Early Colonisation of Land by Fungi and Plants. *Science*, 293:129 - 1133.
- Henk, D., Weir, A. & Blackwell, M., 2003. *Laboulbeniopsis Termitarius*, an Ectoparasite of Termites Newly Recognised as a Member of the *Laboulbeniomyces*. *Mycologia*, 95(4): 561 - 564.
- Hesse, S. J. & Ruijter, G. J., 2002. Intracellular pH Homeostasis in the Filamentous Fungus *Aspergillus niger*. *FEBS*, 269(14): 3485-3494.
- Hibbing, M. E., Fuqua, C., Parsek, M. R. & Peterson, S. B., 2009. Bacterial Competition: Surviving and Thriving in the Microbial Jungle. *Nature Reviews Microbiology*, 8:15-25.
- Hillel, D., 2009. The Mission of Soil Science in a Changing World. *Journal of Plant Nutrition and Soil Science*, 172:5 - 9.
- Hill, G. T. et al. 2000. Methods for Assessing the Composition and Diversity of Soil Microbial Communities. *Applied Soil Ecology*, 15:25-36.
- Hinsinger, P., 2001. Bioavailability of Soil Inorganic P in the Rhizosphere is Affected by Root Induced Chemical Changes: A Review. *Plant and Soil*, 237:173-195.
- Horvath, R. S., 1972. Microbial Co-Metabolism and the Degradation of Organic Compounds in Nature. *Microbiology and Molecular Biology Reviews*, 36(2): 146-155.
- Hugenholtz, P. & Pace, N. R., 1996. Identifying Microbial Diversity in the Natural Environment: a Molecular Phylogenetic Approach. *Trends in Biotechnology*, 14:190 - 197.
- Hugenholtz, P., Tyson, G.W., Webb, R.I., Wagner, A.M., Blackall, L.L., 2001. Investigation of Candidate Division TM7 , a Recently Recognised Major Lineage of the Domain Bacteria with No Known Pure-Culture Representatives. *Applied and Environmental Microbiology*, 67(1):411-419.
- Hull, D., 1999. The Use and Abuse of Sir Karl Popper. *Biology and Philosophy*, 14:481-540.
- Hung, L., 1983. Growth Variation Between and Within Species of Ectomycorrhizal Fungi in Response to pH in Vitro. *Mycologia*, 75(2): 234-241.
- Huse, S. M. et al. 2008. Exploring Microbial Diversity and Taxonomy Using SSU rRNA Hypervariable Tag Sequencing. *PLoS Genetics*, 4.

IGBP-DIS, 1998. SoilData(V.0) A Program for Creating Global Soil-Property Databases. France: IGBP Global Soils Data Task.

Ihaka, R. & Gentleman, R., 1996. R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, 5(3): 299:314.

Ikeda, S. et al. 2008. Evaluation of the Effects of Different Additives in Improving the DNA Extraction Yield and Quality from Andosol. *Microbes and Environments*, 23:159-166.

Isaac, S., 1992. *Fungal-Plant Interactions*. 2nd ed. Cambridge: Chapman & Hall.

Islam, A. K., Edwards, D. G. & Asher, C. J., 1980. pH Optima for Crop Growth. *Plant and Soil*, 54:339-357.

James, T. Y. et al. 2006. A Molecular Phylogeny of the Flagellated Fungi (Chytridiomycota) and Description of a new Phylum (Blastocladiomycota). *Mycologia*, 98(6): 860 - 871.

James, T. Y., Porter D., Leander C. A. Vilgalys, R. & Longcore, J. E. 2000. Molecular Phylogenetics of the Chytridiomycota Supports the Utility of Ultrastructural Data in Chytrid Systematics. *Canadian Journal of Botany*, 78:336 - 350.

Janos, D. P., Sahley, C. & Emmons, L., 1995. Rodent Dispersal of Vesicular-Arbuscular Mycorrhizal Fungi in Amazonian Peru. *Ecology*, 76(6): 1852 - 1858.

Janssen, P., 2006. Identifying the Dominant Soil Bacterial Taxa in Libraries of 16S rRNA and 16S rRNA Genes. *Applied and Environmental Microbiology*, 72(3): 1719-1728.

Janssen, P. H. et al. 2002. Improved Culturability of Soil Bacteria and Isolation in Pure Culture of Novel Members of the Divisions Acidobacteria, Actinobacteria, Proteobacteria, and Verrucomicrobia. *Applied and Environmental Microbiology*, 68(5): 2391 - 2396.

Jenkins, D. et al. 2007. Does Size matter for Dispersal Distance?. *Global Ecology and Biogeography*, 16:415-425.

Jenny, H., 1941. *Factors of Soil Formation*. New York: Dover Publications.

Jenny, H., 1980. *The Soil Resource, Origin and Behaviour*. 1st ed. New York: Springer-Verlag.

Jeong, K. C. et al. 2008. Acid Stress Damage of DNA is Prevented by Dps Binding in Escherichia Coli. *BMC Microbiolgy*, 8:8-181.

Jizhong, Z., Bruns, M. A. & Tiedje, J. M., 1996. DNA Recovery from Soils of Diverse Composition. *Applied and Environmental Microbiology*, 62(2): 316:322.

- Johnson, D. B. & Hallberg, K. B., 2005. Acid Mine Drainage Remediation Options: A Review. *Science of the Total Environment*, 338:3-14.
- Johnson, D. & Worobec, A., 1988. Spatial and Temporal Computer Analysis of Insects and Weather: Grasshoppers and Rainfall in Alberta. *Memoirs of the Entomological Society of Canada*, 120(146): 33-48.
- Johnson, M. J., Lee, K. Y. & Scow, K. M., 2003. DNA Fingerprinting Reveals Links Among Agricultural Crops, Soil Properties, and the Composition of Soil Microbial Communities. *Geoderma*, 114(3): 279-303.
- Johnston, A. E., 2004. Soil Acidity - Resilience and Thresholds. In: P. Schjonning, ed. *Managing Soil Quality: Challenges in Modern Agriculture*. Wallingford: CABI Publishing: 35-44.
- Jones, D. & Muehlchen, A., 1994. Effects of the Potentially Toxic Metals, Aluminium, Zinc and Copper on Ectomycorrhizal Fungi. *Journal of Environmental Science and Health*, 29(5): 949-966.
- Jones, R. T. et al. 2009. A Comprehensive Survey of Soil Acidobacterial Diversity using Pyrosequencing and Clone Library Analysis. *The ISME Journal*, 3:442-453.
- Jongejans, E. & Schippers, P., 1999. Modelling Seed Dispersal by Wind in Herbaceous Species. *Oikos*, 87:362-372.
- Kaiser, C. et al. 2010. Below ground Carbon Allocation by Trees Drives Seasonal Patterns of Extracellular Enzyme Activities by Altering Microbial Community Composition in Beech Forest Soil. *New Phytologist*, 187(3): 834 - 858.
- Keller, D. R. & Golley, F. B., 2000. *The Philosophy of Ecology*. 1st ed. Athens: University of Georgia Press.
- Kielak, A., Pijl, A. S., Veen, J. & Kowalchuk, G., 2009. Phylogenetic Diversity of Acidobacteria in a Former Agricultural Soil. *The ISME Journal*, 3:378-382.
- Kim, M., Morrison, M. & Yu, Z., 2011. Evaluation of Different Partial 16S rRNA Gene Sequence Regions for Phylogenetic Analysis of Microbiomes. *The ISME Journal*, 84(1): 81 - 87.
- Kindt, R. & Coe, R., 2005. *Tree diversity analysis: A manual and software for common statistical and biodiversity methods*. Nairobi: World Agroforestry Centre (ICRAF). ISBN 92-9059-179-X. Available online <http://www.worldagroforestry.org/downloads/publications/PDFs/MN08242.PDF>
- King, A. et al. 2010. Biogeography and Habitat Modelling of High-Alpine Bacteria. *Nature Communications*, 1(53): 1-6.

- Kishimoto, N., Kosako, Y. & Tano, T., 1991. *Acidobacterium capsulatum* gen. nov., sp. nov.: An Acidophilic Chemoautotrophic Bacterium Containing Menaquinone from Acidic Mineral Environment. *Current Microbiology*, 22(1): 1-7.
- Kleupfel, D. A., 1993. The Behaviour and Tracking of Bacteria in The Rhizosphere. *Annual Review of Phytopathology*, 31:441-472.
- Kodie, R. T., Xu, B. & Sharda, J., 2005. Contrasting Below-ground Views of an Ectomycorrhizal Fungal Community. *New Phytologist*, 166:251-262.
- Kourtev, P., Ehrenfeld, J. G. & Haggblom, M., 2002. Exotic Plant Species Alter the Microbial Community Structure and Function in the Soil. *Ecology*, 83:3152-3166.
- Kristjansson, J. K. & Hreggvidsson, G. O., 1995. Ecology and Habitats of Extremophiles. *World Journal of Microbiology and Biotechnology*, 11(1): 17-25.
- Krug, E. C. & Frink, C. R., 1983. Acid Rain on Acid Soil: A New Perspective. *Science*, 221(4610): 520-525.
- Kunin, V. & Hugenholtz, P., 2010. PyroTagger: A fast, accurate pipeline for Analysis of rRNA Amplicon Pyrosequencing Data. *The Open Journal*, 1:1 - 8.
- Kunin, V., Engelbrektson, A., Ochman, H., Hugenholtz, P., 2010. Wrinkles in the Rare Biosphere: Pyrosequencing Errors can Lead to artificial Inflation of Diversity Estimates. *Environmental Microbiology*. 12(1): 118 - 123.
- Kurtzman, C. P., Fell, J. W. & Boekhout, T., 2011. *The Yeasts: A Taxonomic Study*. 5th ed. London: Elsevier.
- Lane, D.J., (1991). 16S/23S rRNA sequencing. In *Nucleic Acid Techniques in Bacterial Systematics*. pp 115 – 175. Edited by E. Stackebrandt & M. Goodfellow. Chichester: Wiley
- Lahlali, R. & Hijri, M., 2010. Screening, Identification and Evaluation of Potential Biocontrol Fungal Endophytes Against *Rhizoctonia Solani* AG3 on Potato Plants. *FEMS Microbiology Letters*, 311(2): 152-159.
- Lauber, C. L., Hamady, M., Knight, R. & Fierer, N., 2009. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Applied and Environmental Microbiology*, 75(15): 5111-5120.
- Lauber, C., Strickland, M., Bradford, M. & Fierer, N., 2008. The Influence of Soil Properties on the Structure of Bacterial and Fungal Communities Across Land-Use Types. *Soil Biology and Biochemistry*, 40:2407 - 2415.
- Legendre, P., 1993. Spatial Autocorrelation: Trouble or new Paradigm?. *Ecology*, 74(6): 1659-1673.

- Legendre, P., 2000. Comparison of Permutation Methods for the Partial Correlation and Partial Mantel Tests. *Journal of Statistical Computation and Simulation*, 67(1): 37-73.
- Legendre, P. & Fortin, M., 1989. Spatial Patterns and Ecological Analysis. *Vegetatio*, 80: 107-138.
- Leininger, S. et al. 2006. Archaea Predominate Among Ammonia-Oxidizing Prokaryotes in Soils. *Nature*, 442(17): 806-809.
- Leiros, M. C., Trasar-Cepeda, C., Seoane, S. & Gil-Sotres, F., 1999. Dependence of Mineralization of Soil Organic Matter on Temperature and Moisture. *Soil Biology and Biochemistry*, 31(3): 327-335.
- Lentendu, G. et al. 2011. Assessment of Soil Fungal Diversity in Different Alpine Tundra Habitats by Means of Pyrosequencing. *Fungal Diversity*, 49(1): 113-123.
- Letunic, I. & Bork, P., 2011. Interactive Tree of Life v2: Online Annotation and Display of Phylogenetic Trees made Easy. *Nucleic Acids Research*, 39:475 - 478.
- Leveau, J. H. & Preston, G. M., 2007. Bacterial Mycophagy: Definition and Diagnosis of a Unique Bacterial-Fungal Interaction. *New Phytologist*, 177:859-876.
- Lewen-Koh, N. J. & Bivand, R., 2011. Maptools: Tools for Reading and Handling Spatial Objects. [Online]
- Available at: <http://CRAN.R-project.org/package=maptools>
- Liesack, W., Bak, F., Kreft, J. & Stackbrandt, E., 1994. *Holophaga foetida* gen. nov., sp. nov., a new, Homoacetogenic Bacterium Degrading Methoxylated Aromatic Compounds. *Archives of Microbiology*, 162:85-90.
- Liggenstoffer, A., Youssef, N., Couger, M. & Elshahed, M., 2010. Phylogenetic Diversity and Community Structure of Anerobic Gut Fungi (Phylum Neocallimastigomycota) in Ruminant and Non - Ruminant Herbivores. *The ISME Journal*, 4:1225 - 1235.
- Lilleskov, E. A. et al. 2004. Detection of Forest Stand-level Spatial Structure in Ectomycorrhizal Fungal Communities. *FEMS Microbiology Ecology*, 49:319-332.
- Lim, Y. et al. 2010. Assessment of Soil Fungal Communities using Pyrosequencing. *Journal of Microbiology*, 48(3): 284 - 289.
- Lipke, P. N. & Ovalle, R., 1998. Cell Wall Architecture in Yeast: New Structure and New Challenges. *Journal of Bacteriology*, 180(15): 3735-3740.

Liu, W.-T., Marsh, T., Cheng, H. & Forney, L., 1997. Characterization of Microbial Diversity by Determining Terminal Restriction Fragment Length Polymorphisms of Genes Encoding 16S rRNA. *Applied and Environmental Microbiology*, 63(11): 4516:4522.

Liu, Z., DeSantis, T. Z., Andersen, G. L. & Knight, R., 2008. Accurate Taxonomy Assignments from 16S rRNA Sequences Produced by Highly Parallel Pyrosequencers. *Nucleic Acids Research*, 36(18): e120.

Liu, Z. et al. 2007. Short Pyrosequencing Reads Suffice for Accurate Microbial Community Analysis. *Nucleic Acids Research*, 35(18): e120 .

Longnecker, K. & Reysenbach, A., 2001. Expansion of the Geographic Distribution of a Novel Lineage of Epsilon Proteobacteria to a Hydrothermal Vent site on the Southern East Pacific Rise. *FEMS Microbiology Ecology*, 35: 287 - 293.

Lord, N. S., Kaplan, C. W., Shank, P. & Kitts, C. L., 2002. Assessment of Fungal Diversity Using Terminal Restriction Fragment (TRF) Pattern Analysis: Comparison of 18S and ITS ribosomal regions. *FEMS Microbiology Ecology*, 42(3): 327-337.

Lozupone, C. A., Hamandy, M., Kelly, S. T. & Knight, R., 2007. Quantitative and Qualitative Beta Diversity Measures Lead to Different Insights into Factors that Structure Microbial Communities. *Applied and Environmental Microbiology*, 73:1576-1585.

Lozupone, C. A. & Knight, R., 2007. Global patterns in bacterial diversity. *PNAS*, 104(27): 11436-11440.

Lozupone, C. & Knight, R., 2005. UniFrac: A New Phylogenetic Method for Comparing Microbial Communities. *Applied and Environmental Microbiology*, 71(12): 8228 - 8235.

Lozupone, C. et al. 2011. UniFrac: an effective distance metric for microbial community comparison. *The ISME Journal*, 5: 169 - 172.

Luders, T. & Friedrich, M. W., 2003. Evaluation of PCR Amplification Bias by Terminal Restriction Fragment Length Polymorphism Analysis of Small-Subunit rRNA and mcrA Genes by Using Defined Template Mixtures of Methanogenic Pure Cultures and Soil DNA Extracts. *Applied and Environmental Microbiology*, 69(1): 320-326.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., Forster, W., Brettske, I., Gerber, S., Ginhart, A.W., Gross, O., Grumann, S., Hermann, S., Jost, R., Koing, A., Liss, T., Lussmann, R., May, M., Nonhoff, B., Reichel, B., Strehlow, R., Stamatakis, A., Stuckmann, N., Vilbig, A., Lenke, M., Ludwig, T., Bode, A., Schleifer, K.H.,

2004. ARB: A Software Environment for Sequence Data. *Nucleic Acids Research*, 32(4): 1363-1371.
- Lupwayi, N. Z., Rice, W. A. & Clayton, G. W., 1998. Soil Microbial Diversity and Community Structure Under Wheat as Influenced by Tillage and Crop Rotation. *Soil Biology and Biochemistry*, 30(13): 1733-1741.
- MacArthur, R. H., 1972. *Geographical Ecology: Patterns in the Distribution of Species*. Princetown: Princetown University Press.
- Madigan, M. T., Martinko, J. M. & Parker, J., 2003. *Brock Biology of Microorganisms*. 18th ed. Upper Saddle River: Hall/Pearson Education.
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., 2011. *Cluster: Cluster Analysis Basics and Extensions*. R package version 1.14.2.
- Mainfield, M., Whiteley, A. S., Griffiths, R. I. & Bailey, M. J., 2002. RNA Stable Isotope Probing, a Novel Means of Linking Microbial Community Function to Phylogeny. *Applied and Environmental Microbiology*, 68(11): 5367-5373.
- Malik, M., Kain, J., Pettigrew, C. & Ogram, A., 1994. Purification and Molecular Analysis of Microbial DNA from Compost. *Journal of Microbiological Methods*, 20(3): 183-196.
- Malloch, D. W., Pirozynski, K. A. & Raven, P. H., 1980. Ecological and Evolutionary Significance of Mycorrhizal Symbioses in Vascular Plants (A Review). *PNAS*, 77(4): 2113-2118.
- Mankau, R., 1980. Biological Control of Nematode Pests by Natural Enemies. *Phytopathology*, 18:415-440.
- Mantel, N., 1967. The Detection of Disease Clustering and a Generalized Regression Approach. *Cancer Research*, 27(2): 209-220.
- Marinari, S., Masciandaro, G., Ceccanti, B. & Grego, S., 2000. Influence of Organic and Mineral Fertilisers on Soil Biological and Physical Properties. *Bioresource Technology*, 72(1): 9-17.
- Marschner, P., Yang, C. H., Lieberei, R. & Crowley, D. E., 2001. Soil and Plant Specific Effects on Bacterial Community Composition in the Rhizosphere. *Soil Biology and Biochemistry*, 33(11): 1437-1445.
- Marsh, T. L., 1999. Terminal Restriction Fragment Length Polymorphism (T-RFLP): An Emerging Method for Characterizing Diversity Among Homologous Populations of Amplification Products. *Current Opinion in Microbiology*, 2(3): 323-327.
- Marchesi, J., Sato, T., Weightman, A., Martin, T., Fry, J., Hiom, S., Wade, W., 1998. Design and Evaluation of Useful Bacterium-Specific PCR Primers That

Amplify Genes Coding for Bacterial 16S rRNA. *Applied and Environmental Microbiology* 64(2), pp 795 – 799.

Martiny, J., Bohannan, B., Brown, J., Colwell, R., Furman, J., Green, J., Devine, M., Kane, M., Krummins, J., Kuske, C., Staley, J., 2006. Microbial Biogeography: Putting Microorganisms on the Map. *Nature Reviews: Microbiology*, 4:102-112.

Matheny, P.B., Curtis, J.M., Hofstetter, V., Aime, M.C., Moncalvo, J.M., Slot, J.C., Ammirati, J.F., Baroni, T.J., Bougher, N.L., Hughes, K.W., Lodge, D.J., Kerrigan, R.W., Seidl, M.T., Aanen, D.K., DeNitis, M., Daniele, G.M., Desjardin, D.E., Kropp, B.R., Norvell, L.L., Parker, A., Vellinga, E.C., Vilgalys, R., Hibbett, D.S. 2006. Major Clades of the Agaricales: a Multilocus Phylogenetic Overview. *Mycologia*, 98:984 - 997.

Metzker, M. L., 2010. Sequencing Technologies - The Next Generation. *Nature Reviews Genetics*, 11:31 - 46.

Millanes, A., Diederich, P., Ekman, S. & Wedin, M., 2011. Phylogeny and Character Evolution in the Jelly Fungi (Tremellomycetes, Basidiomycota, Fungi). *Molecular Phylogenetics and Evolution*, 61(1): 12 - 28.

Miller, D. N., Bryant, J. E., Madsen, E. L. & Ghorse, W. C., 1999. Evaluation and Optimisation of DNA Extraction and Purification Procedures for Soil and Sediment Samples. *Applied and Environmental Microbiology*, 65(11): 4715-4724.

Miller, H. J., 2004. Tobler's First Law and Spatial Analysis. *Annals of the Association of American Geographers*, 94(2): 284-289.

Minchin, P. R., 1987. An Evaluation of the Relative Robustness of Techniques for Ecological Ordination. *Vegetatio*, 69: 89-107.

Mitchell, R. J. et al. 2011. Explaining the Variation in the Soil Microbial Community: Do Vegetation Composition and Soil Chemistry Explain the Same or Different Parts of the Microbial Variation?. *Plant Soil*, p. Online First.

Mittelbach, G. G., 2007. Evolution and the Latitudinal Diversity Gradient: Speciation, Extinction and Biogeography. *Ecology Letters*, 10(4): 315-331.

Montanarella, L., Jones, R. J. & Hiederer, R., 2006. The Distribution of Peatland in Europe. *Mires and Peat*, 1(1): 1-10.

Montealegre, C. M., Kessel, C., Russelle, M. & Sadowsky, M. J., 2002. Changes in Microbial Activity and Composition in a Pasture Ecosystem Exposed to Elevated Atmospheric Carbon Dioxide. *Plant and Soil*, 243(2): 197-207.

Moora, M. et al. 2011. Alien Plants Associate with Widespread Generalist Arbuscular Mycorrhizal Fungal Taxa: Evidence from a Continental Scale Study

Using Massively Parallel 454 Sequencing. *Journal of Biogeography*, 38(7): 1305 - 1317.

Morgan, J. A. W., Bending, G. D. & White, P. J., 2005. Biological Costs and Benefits to Plant-Microbe Interactions in the Rhizosphere. *Journal of Experimental Botany*, 56(417): 1729-1739.

Morgenstern, I., Klopman, S. & Hibbett, D. S., 2008. Molecular Evolution and Diversity of Lignin Degrading Heme Peroxidases in the Agaricomycetes. *Molecular Evolution*, 66(3): 243-257.

Mueller, L. et al. 2010. Assessing the productivity function of soils. A review. *Agronomy for Sustainable Development*, 30: 601-614.

Mulder, C. et al. 2005. Bacterial Traits, Organism Mass, and Numerical Abundance in the Soil Food Web of Dutch Agricultural Grasslands. *Ecology Letters*, 8(1): 80-90.

Muller, A. K., Westergaard, K., Christensen, S. & Sorensen, S. J., 2001. The Effect of Long Term Mercury Pollution on the Soil Microbial Community. *FEMS Microbiology Ecology*, 36(1): 11-19.

Mummey, D. L., Rillig, M. C. & Holben, W. E., 2005. Neighbouring plant influences on arbuscular mycorrhizal fungal community composition as assessed by T-RFLP analysis. *Plant and Soil*, 271(1-2): 83-90.

Mummey, D. L. & Rilling, M. C., 2006. The Invasive Plant Species *Centaurea maculosa* Alters Arbuscular Mycorrhizal Fungal Communities in the Field. *Plant and Soil*, 288:81-90.

Muyzer, G. & Smalla, K., 1998. Application of Denaturing Gradient Gel Electrophoresis (DGGE) and Temperature Gradient Gel Electrophoresis (TGGE) in Microbial Ecology. *Antonie Van Leeuwenhoek*, 73(1): 127:141.

Nacke, H. et al. 2011. Pyrosequencing-Based Assessment of Bacterial Community Structure Along Different Management Types in German Forest and Grassland Soils. *PLoS One*, 16(2): 1-12.

Nakatsu, C. H., Torsvik, V. & Øvreås, L., 2000. Soil Community Analysis using DGGE of 16S rDNA Polymerase Chain Reaction Products. *Soil Science Society of America Journal*, 64(4): 1382 - 1388.

Nannipieri, P. et al. 2003. Microbial Diversity and Soil Functions. *European Journal of Soil Science*, 54:655-670.

Nardi, S. et al. 2000. Soil Organic Matter Mobilization by Root Exudates. *Chemosphere*, 41(5): 653-658.

- Neefs, J., Peer, Y., Hendriks, L. & Wachter, R., 1990. Compilation of Small Ribosomal Subunit RNA Sequences. *Nucleic Acids Research*, 18(Supplement): 2237 - 2317.
- Nicol, G., Leininger, S., Schleper, C. & Prosser, J., 2008. The Influence of Soil pH on the Diversity, Abundance and Transcriptional Activity of Ammonia Oxidizing Archaea and Bacteria. *Environmental Microbiology*, 10(11): 2966 - 2978.
- Nilsson, H. et al. 2009. A Software Pipeline for Processing and Identification of Fungal ITS Sequences. *Source Code for Biology and Medicine*, 4: 1.
- Nilsson, H. et al. 2009. The ITS Region as a Target for Characterization of Fungal Communities using Emerging Sequence Technologies. *FEMS Microbiology Letters*, 296(1): 97-101.
- Nilsson, R. H. et al. 2008. Intraspecific ITS Variability in the Kingdom Fungi as Expressed in the International Sequence Databases and Implications for Molecular Species Identification. *Evolutionary Bioinformatics*, 4:193-201.
- Nilsson, R. H. et al. 2006. Taxonomic Reliability of DNA Sequences in Public Sequence Databases: A Fungal Perspective. *PLoS ONE*, 1(1): e59.
- Noller, H. F. & Woese, C. R., 1981. Secondary Structure of 16S ribosomal RNA. *Science*, 212: 403 - 411.
- Norris, T. B., Wraith, J. M., Castenholz, R. W. & McDermott, T. R., 2002. Soil Microbial Community Structure across a Thermal Gradient following a Geothermal Heating Event. *Applied and Environmental Microbiology*, 68(12): 6300-6309.
- Norton, J. M. & Stark, J. M., 2011. Regulation and Measurement of Nitrification in Terrestrial Systems. In: M. G. Klotz, ed. *Research on Nitrification and Related Processes: Part 1*. London: Elsevier: 343-368.
- Nunan, N. et al. 2002. In Situ Spatial Patterns of Soil Bacterial Populations, Mapped at Multiple Scales, in an Arable Soil. *Microbial Ecology*, 44: 296-305.
- Nunan, N. et al. 2003. Spatial Distribution of Bacterial Communities and their Relationships with the Micro-Architecture of soil. *FEMS Microbial Ecology*, 44: 203-215.
- Nwoke, O. C. et al. 2003. Assessments of Labile Phosphorus Fractions and Adsorption Characteristics in Relation to Soil Properties of West African Savannah Soils. *Agriculture, Ecosystems and Environment*, 100(2-3): 285-294.
- Odum, E. & Barrett, G., 2004. *Fundamentals of Ecology*. 5 ed. California: Brooks Cole.

- Eriksson, O.E., & Hawksworth, D.L., 2005. Outline of the Ascomycota. *Myconet*, 11: 1 - 113.
- Ogram, A., Sayler, G. S. & Barkay, T., 1987. The Extraction and Purification of Microbial DNA from Sediments. *Journal of Microbiological Methods*, 7(2-3): 57-66.
- Oksanen, J., 2011. Multivariate Analysis of Ecological Communities in R: Vegan Tutorial. [Online]
- Available at: <http://cc.oulu.fi/~jarioksa/opetus/metodi/vegantutor.pdf>
- [Accessed 10 1 2010].
- Oksanen, J. et al. 2011. Vegan: Community Ecology Package.
- O'Malley, M. A., 2007. The Nineteenth Century Roots of "everything is everywhere". *Nature Reviews: Microbiology*, 5:647 - 651.
- O'Toole, G., Kaplan, H. B. & Kolter, R., 2000. Biofilm Formation as Microbial Development. *Annual Review of Microbiology*, 54:49-79.
- Pan, J. & Yu, L., 2011. Effects of Cd or/and Pb on Soil Enzyme Activities and Microbial Community Structure. *Ecological Engineering*, 37(11): 1889-1894.
- Papke, T. R., Ramsing, N. B., Bateson, M. M. & Ward, D. M., 2003. Geographical Isolation in Hot Spring Cyanobacteria. *Environmental Microbiology*, 5(8): 650-659.
- Partel, M., 2002. Local Plant Diversity Patterns and Evolutionary History at the Regional Scale. *Ecology*, 83(9): 2361 - 2366.
- Peakall, R., Ruibal, M. & Lindenmayer, D., 2003. Spatial Autocorrelation Analysis Offers new Insights into Gene flow in the Australian Bush Rat, *Rattus fuscipes*. *Evolution*, 57(5): 1182-1195.
- Peay, K. G. et al. 2007. A Strong Species-Area Relationship for Eukaryotic Soil Microbes: Island Size Matters for Ectomycorrhizal Fungi. *Ecology Letters*, 10:470-480.
- Pebesma, E. J., 2004. Multivariable geostatistics in S: the gstat package. *Computers and Geosciences*, 30:683-691.
- Peel, M. C., Finlayson, B. L. & McMahon, T. A., 2007. Updated World Map of the Koppen-Geiger Climate Classification. *Hydrology and Earth System Sciences*, 11:1633-1644.
- Pester, M. et al. 2010. A "Rare Biosphere" Microorganism Contributes to Sulphate Reduction in a Peatland. *The ISME Journal*, 4:1591 - 1602.

- Picard, C. et al. 1992. Detection and Enumeration of Bacteria in Soil by Direct DNA Extraction and Polymerase Chain Reaction. *Applied and Environmental Microbiology*, 58(9): 2713-2722.
- Pielou, E.C., 1966: The measurement of diversity in different types of biological collections. *Journal of Theoretical Biology* 13: 131–144.
- Pietri, A. J. & Brookes, P. C., 2008. Relationships between Soil pH and Microbial Properties in a UK Arable Soil. *Soil Biology and Biochemistry*, 40: 1856-1861.
- Pina, R. G. & Cervantes, C., 1996. Microbial Interactions with Aluminium. *Biometals*, 9:311-316.
- Pirozynski, K. A. & Malloch, D. W., 1975. The Origin of Land Plants: A Matter of Mycotrophism. *Biosystems*, 6(3): 153-164.
- Pitcher, D. G., Saunders, N. A. & Owen, R. J., 1989. Rapid Extraction of Bacterial Genomic DNA with Guanidium Thiocyanate. *Letters in Applied Microbiology*, 8(4): 151-156.
- Poll, C., Brune, T. & Begerow, D., 2010. Small-scale Diversity and Succession of Fungi in the Detritosphere of Rye Residues. *Microbial Ecology*, 59(1): 130-140.
- Popper, K. P., 1959. *The Logic of Scientific Discovery*. 1st ed. London: Hutchinson & Co.
- Porteous, A. L. & Armstrong, J. L., 1991. Recovery of bulk DNA from soil by a rapid, small-scale extraction method. *Current Microbiology*, 22(6): 345-348.
- Post, W. M., Emanuel, W. R., Zinke, P. J. & Stagenberger, A. G., 1982. Soil Carbon Pools and World Life Zones. *Nature*, 298: 156-159.
- Price, M. N., Dehal, P. S. & Arkin, A. P., 2010. FastTree 2 - Approximately Maximum Likelihood Trees for Large Alignments. *PLoS ONE*, 5(3): e9490.
- Pruesse, E. et al. 2007. SILVA: a Comprehensive Online Resource for Quality Checked and Aligned Ribosomal RNA Sequence Data Compatible with ARB. *Nucleic Acids Research*, 35(21): 7188 - 7196.
- Pryde, S. E., Richardson, A. J., Stewart, C. S. & Flint, H. J., 1999. Molecular Analysis of the Microbial Diversity Present in the Colonic Wall, Colonic Lumen and Cecal Lumen of Pigs. *Applied and Environmental Microbiology*, 65(12): 5372-5377.
- Quince, C. et al. 2009. Accurate Determination of Microbial Diversity from 454 Pyrosequencing Data. *Nature Methods*, 6:639-641.

- Ramette, A. & Tiedje, J. M., 2007. Biogeography: An Emerging Cornerstone for Understanding Prokaryote Diversity, Ecology, and Evolution. *Microbial Ecology*, 53:197 - 207.
- Ranjard, L. et al. 2011. Biogeography of Soil Microbial Communities: A Review and Description of the On-going French National Initiative. *Sustainable Agriculture Two*, 2(6): 857-865.
- Ranjard, L., Poly, F. & Nazaret, S., 2000. Monitoring Complex Bacterial Communities Using Culture-Independent Molecular Techniques: Application to Soil Environment. *Research in Microbiology*, 151(3): 167-177.
- Ranjard, L. & Richaume, A., 2001. Quantitative and Qualitative Microscale Distribution of Bacteria in Soil. *Research in Microbiology*, 152(8): 707-716.
- Raymond, V. B., O'Connor, J. O. & Krohn, W. B., 2006. Importance of Spatial Autocorrelation in Modelling Bird Distributions at a Continental Scale. *Ecography*, 29:835-844.
- Read, D. J. et al. 2000. Symbiotic Fungal Associations in "Lower" Land Plants. *Philosophical Transactions of the Royal Society B Biological Sciences*, 355(1398): 815-831.
- Reddy, G. S. N., & Garcia-Pichel, F., 2009. Description of *Patulibacter americanus* sp. nov., Isolated from Biological Soil Crusts, Emended Description of the Genus *Patulibacter* Takahashi et al. 2006 and Proposal of *Solirubrobacterales* ord. nov. and *Thermoleophilales* ord. nov.. *IJSEM*, 59: 87 - 94.
- Rengel, Z., 2003. *Handbook of Soil Acidity*. Ney York: Marcel Dekker.
- Rheims, H., Rainey, F.A., Stackebrandt, E., 1996. A Molecular Approach to Search for Diversity Among Bacteria in the Environment. *Journal of Industrial Microbiology*, 17, 159-169.
- Ricke, P., Kolb, S. & Braker, G., 2005. Application of a Newly Developed ARB Software-Integrated Tool for In Silico Terminal Restriction Fragment Length Polymorphism Analysis Reveals the Dominance of a Novel *pmoA* Cluster in a Forest Soil. *Applied and Environmental Microbiology*, 71(3): 1671-1673.
- Rinaldi, A. C., Comandini, O. & Kuyper, T. W., 2008. Ectomycorrhizal Fungal Diversity: Separating the Wheat from the Chaff. *Fungal Diversity*, 33:1-45.
- Ritz, K. & Young, I., 2004. Interactions Between Soil Structure and Fungi. *Mycologist*, 18(2): 52-59.
- Roberts, D.W., (2010). LabDSV: Ordination and Multivariate Analysis for Ecology. R Package Version 1.4-1. [HTTP://CRAN.R-project.org/package=labdsv](http://CRAN.R-project.org/package=labdsv)

- Roesch, L. F. et al. 2007. Pyrosequencing Enumerates and Contrasts Soil Microbial Diversity. *The ISME Journal*, 1:283 - 290.
- Rondon, M. R., Goodman, R. M. & Handelsman, J., 1999. The Earth's bounty: assessing and accessing soil microbial diversity. *Trends in Biotechnology*, 17(10): 403-409.
- Ronn, R., McCaig, A. E., Griffiths, B. S. & Prosser, J. I., 2002. Impact of Protozoan Grazing on Bacterial Community Structure in Soil Microcosms. *Applied and Environmental Microbiology*, 68(12): 6094-6105.
- Rooney, A. & Ward, T. J., 2005. Evolution of a Large Ribosomal RNA Multigene Family in Filamentous Fungi: Birth and Death of a Concerted Evolution Paradigm. *PNAS*, 102(14): 5084-5089.
- Roose-Amsaleg, C. L., Garnier-Sillam, E. & Harry, M., 2001. Extraction and purification of microbial DNA from soil and sediment samples. *Applied Soil Ecology*, 18(1): 47:60.
- Roseburg, R. J., Christensen, N. W. & Jackson, T. L., 1985. Chloride, Soil Solution Osmotic Potential and Soil pH Effects on Nitrification. *Soil Science Society of America Journal*, 50(4): 941-945.
- Rouatt, J. W. & Katznelson, H., 1961. A Study of the Bacteria on the Root Surface and in the Rhizosphere of Crop Plants. *Applied Microbiology*, 24(2): 164-171.
- Rousk, J. et al. 2010. Soil Bacterial and Fungal Communities Across a pH Gradient in an Arable Soil. *The ISME Journal*, 4:1340-1351.
- Rousk, J., Brookes, P. C. & Baath, E., 2009. Contrasting Soil pH Effects on Fungal and Bacterial Growth Suggest Functional Redundancy in Carbon Mineralization. *Applied and Environmental Microbiology*, 75:1589 - 1596.
- Ruibal, C. et al. 2009. Phylogeny of Rock-Inhabiting Fungi Related to the Dothideomycetes. *Studies in Mycology*, 64(1): 123 - 133.
- Sait, M., Davis, K. E. & Janssen, P. H., 2006. Effect of pH on Isolation and Distribution of Members of Subdivision 1 of the Phylum Acidobacteria Occurring in Soil. *Applied and Environmental Microbiology*, 72(3): 1852-1857.
- Sait, M., Hugenholtz, P. & Janssen, P., 2002. Cultivation of Globally Distributed Soil Bacteria from Phylogenetic Lineages Previously only Detected in Culture-Independent Surveys. *Environmental Microbiology*, 4(11): 654 - 666.
- Salman, V. et al. 2011. A Single-Cell Sequencing Approach to the Classification of Large, Vacuolated Sulphur Bacteria. *Systematic and Applied Microbiology*, 34(4): 243-259.

- Schlesinger, W. H., 1977. Carbon Balance in Terrestrial Detritus. *Annual Review of Ecology and Systematics*, 8:51-81.
- Schloss, P. D., 2010. The Effects of Alignment Quality, Distance Calculation Method, Sequence Filtering, and Region on the Analysis of 16S rRNA Gene-Based Studies. *PLoS Computational Biology*, 6(7): e1000844. doi:10.1371/journal.pcbi.1000844.
- Schloss, P. D. & Handelsman, J., 2006. Towards a Census of Bacteria in Soil. *PLoS Computational Biology*, 2(7): e92. doi:10.1371/journal.pcbi.0020092.
- Schüßler, A., Schwarzott, D. & Walker, C., 2001. A New Fungal Phylum, the Glomeromycota: Phylogeny and Evolution. *Mycological Research*, 105(12): 1413 - 1421.
- Seifert, K., 2009. Progress towards DNA barcoding of Fungi. *Molecular Ecology Resources*, 9(S1): 83 - 89.
- Sessitsch, A. et al. 2001. Microbial Population Structures in Soil Particle Size Fractions of a Long Term Fertilizer Field Experiment. *Applied and Environmental Microbiology*, 67(9): 4215-4224.
- Singleton, D. R. et al. 2003. *Solirubrobacter pauli* gen. nov., sp. nov., a Mesophilic Bacterium Within the Rubrobacteridae Related to Common Soil Clones. *IJSEM*, 45: 485 - 490.
- Sletten, O. & Skinner, C., 1948. Fungi Capable of Growing in Strongly Acid Media and Concentrated Copper Sulphate solutions. *Journal of Bacteriology*, 56(5): 679-681.
- Smalla, K. et al. 2008. Rapid DNA Extraction Protocol from Soil for Polymerase Chain Reaction-Mediated Amplification. *Journal of Applied Bacteriology*, 74(1): 78:85.
- Smit, E. et al. 1999. Analysis of Fungal Diversity in the Wheat Rhizosphere by Sequencing of Cloned PCR-Amplified Genes Encoding 18S rRNA and Temperature Gradient Gel Electrophoresis. *Applied and Environmental Microbiology*, 66(6): 2614-2621.
- Smit, E. et al. 2001. Diversity and Seasonal Fluctuations of the Dominant Members of the Bacterial Soil Community in a Wheat Field as Determined by Cultivation and Molecular Methods. *Applied and Environmental Microbiology*, 67(5): 2284-2291.
- Smit, E., Leeflang, P. & Wernars, K., 1997. Detection of Shifts in Microbial Community Structure and Diversity in Soil Caused by Copper Contamination using Amplified Ribosomal DNA Restriction Analysis. *FEMS Microbial Ecology*, 23(3): 249-261.

- Smith, J. M., Dowson, C. G. & Spratt, B. G., 1991. Localized Sex in Bacteria. *Nature*, 349(6304): 29 - 31.
- Spain, A. M., Krumholz, L. R. & Elshahed, M. S., 2009. Abundance, Composition, Diversity and Novelty of Soil Proteobacteria. *The ISME Journal*, 3: 992-1000.
- Spatafora, J. et al. 2006. A Five-Gene Phylogeny of the Pezizomycotina. *Mycologia*, 98(6): 1018 - 1028.
- Stach, J. E., Bathe, S., Clapp, J. P. & Burns, R. G., 2006. PCR-SSCP Comparison of 16S rDNA Sequence Diversity in Soil DNA Obtained Using Different Isolation and Purification Methods. *FEMS Microbiology Ecology*, 36:139-151.
- Stackebrandt, E., Murray, R. G. & Truper, H. G., 1988. Proteobacteria classis nov., a Name for the Phylogenetic Taxon that Includes the "Purple Bacteria and their Relatives". *International Journal of Systematic Bacteriology*, 38(3): 321-325.
- Stackebrandt, E., Rainey, F. A. & Ward-Rainey, N., 1997. Proposal for a New Hierarchic Classification System, Actinobacteria classis nov.. *IJSEM*, 47(2): 479 - 491.
- Stark, N., 1972. Nutrient Cycling Pathways and Litter Fungi. *Bioscience*, 22(6): 355-360.
- Ste-Marie, C. & Pare, D., 1999. Soil, pH and N Availability effects on net Nitrification in the Forest Floors of a Range of Boreal Forest Stands. *Soil Biology and Biochemistry*, 31(11): 1579-1589.
- Stockmann, U., Minasny, B. & McBratney, A., 2011. Quantifying Processes of Pedogenesis. *Advances in Agronomy* 113: 1-72.
- Strickland, M. S. et al. 2009. Litter Quality is in the Eye of the Beholder: Initial Decomposition Rates as a Function of Innoculum Characteristics. *Functional Ecology*, 90:441-451.
- Suzuki, M. & Giovannoni, S. J., 1996. Bias Caused by Template Annealing in the Amplification of Mixtures of 16S rRNA Genes by PCR. *Applied and Environmental Microbiology*, 62(2): 625-630.
- Sylvia, D. M., Fuhrmann, J. F., Hartel, P. G. & Zuberer, D. A., 2005. Principles and Applications of Soil Microbiology. New Jersey: Pearson Education Inc.
- Takada-Hoshino, Y. & Matsumoto, N., 2004. An Improved DNA Extraction Method Using Skim Milk from Soils That Strongly Adsorb DNA. *Microbes and Environments*, 19(1): 13-19.

Tarlera, S. et al. 2008. Microbial Community Succession and Bacterial Diversity in Soils During 77 000 Years of Ecosystem Development. *FEMS Microbiology Ecology*, 64(1): 129-140.

Taylor, J. P., Wilson, B., Mills, M. S. & Burns, R. G., 2002. Comparison of Microbial Numbers and Enzymatic Activities in Surface Soils and Subsoils using Various Techniques. *Soil Biology and Biochemistry*, 34(3): 387-401.

Tebbe, C. C. & Vahjen, W., 1993. Interference of Humic Acids and DNA Extracted Directly from Soil in Detection and Transformation of Recombinant DNA from Bacteria and a Yeast. *Applied and Environmental Microbiology*, 59(8): 2657:2665.

Thomas, G. W., 1996. Soil pH and Soil Acidity. In: D. Sparks, ed. *Methods of Soil Analysis. Part 3 - Chemical Methods..* Wisconsin: Soil Science Society of America:475-490.

Thompson, J. D., Higgins, D. G. & Gibson, T. J., 1994. CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment Through Sequence Weighting, Position Specific Gap Penalties and Weight Matrix Choice. *Nucleic Acids Research*, 22:4673 - 4680.

Thomson, B. C. et al. 2010. Effects of Sieving, Drying and Rewetting upon Soil Bacterial Community Structure and Respiration Rates. *Journal of Microbiological Methods*, 83(1): 69:73.

Thomson, B. C. et al. 2010. Vegetation Affects the Relative Abundances of Dominant Soil Bacterial Taxa and Soil Respiration Rates in an Upland Grassland Soil. *Microbial Ecology*, 59: 335 - 343.

Tiedje, J. M. et al. 1998. Opening the Black Box of Soil Microbial Diversity. *Applied Soil Ecology*, 13(2): 109:122.

Tiedje, J. M. et al. 2001. Soil Teeming With Life: New Frontiers for Soil Science. In: *Sustainable Management of Soil Organic Matter*. Wallingford: CAB International: 393-412.

Timmis, K. N. & Pieper, D. H., 1999. Bacteria Designed for Bioremediation. *Trends in Biotechnology*, 17(5): 201-204.

Tiquia, S. M. et al. 2002. Effects of Mulching and Fertilization on Soil Nutrients, Microbial Activity and Rhizosphere Bacterial Community Structure determined by Analysis of TRFLPs of PCR-Amplified 16S rRNA genes. *Applied Soil Ecology*, 21(1): 31-48.

Torsvik, V., Goksøyr, J. & Daae, F. L., 1990. High Diversity in DNA of Soil Bacteria. *Applied and Environmental Microbiology*, 56(3): 782-787.

Torsvik, V. & Ovreas, L., 2002. Microbial Diversity and Function in Soil: From Genes to Ecosystems. *Current Opinion in Microbiology*, 5: 240-245.

Tristan, A. et al. 2007. Global Distribution of Pantone-Valentine Leukocidin-Positive Methicillin-Resistant *Staphylococcus aureus*, 2006. *Emerging Infectious Diseases*, 13(4): 594-600.

Tsai, Y.-L. & Olson, B. H., 1992. Rapid Method for Separation of Bacterial DNA from Humic Substances in Sediments for Polymerase Chain Reaction. *Applied and Environmental Microbiology*, 58(7): 2292-2295.

Tsui, K. M. et al. 1998. The Effect of Human Disturbance on Fungal Diversity in the Tropics. *Fungal Diversity*, 1: 19 - 26.

Urban, D. L., 2003. Spatial Analysis in Ecology. [Online]

Available at:
<http://www.nceas.ucsb.edu/files/scicomp/Dloads/SpatialAnalysisEcologists/SpatialEcologyMantelTest.pdf>

[Accessed 10 1 2012].

Van Dijk, T., 2008. *Microbial Ecology Research Trends*, 1st Edition, New York, Nova Science Publishers Inc. Google Books. Web. 8th August 2010.

van der Heijden, M., Bardgett, R. & van Straalen, N., 2008. The Unseen Majority: Soil Microbes as Drivers of Plant Diversity and Productivity in Terrestrial Ecosystems. *Ecology Letters*, 11: 296 - 310.

Van Ooij, C., 2011. Symbiosis: Fungus seeks plant. *Nature Reviews Microbiology*, 9:148-149.

Venables, W. N. & Ripley, B. D., 2002. *Modern Applied Statistics with S*. New York: Springer.

Ventura, M. et al. 2007. Genomics of Actinobacteria: Tracing the Evolutionary History of an Ancient Phylum. *Microbiology and Molecular Biology Reviews*, 71(3): 495-548.

Vickery, J. et al. 2001. The Management of Lowland Neutral Grasslands in Britain: Effects of Agricultural Practices on Birds and their Food Resources. *Journal of Applied Ecology*, 38: 647 - 664.

Volossiouk, T., Robb, E. J. & Nazar, R. N., 1995. Direct DNA Extraction for PCR-Mediated Assays of Soil Organisms. *Applied and Environmental Microbiology*, 61(11): 3972-3976.

Waksman, S. & Woodruff, B., 1940. Bacteriostatic and Bactericidal Substances Produced by Soil Actinomyces. *Experimental Biology and Medicine*, 45(2): 609-614.

- Walker, T. S., Bais, H. P., Grotewold, E. & Vivanco, J. M., 2003. Root Exudation and Rhizosphere Biology. *Plant Physiology*, 132: 44-51.
- Wang, Q., Garrity, G. M., Tiedge, J. M. & Cole, J. R., 2007. Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the new Bacterial Taxonomy. *Applied and Environmental Microbiology*, 73(16): 5261 - 5267.
- Wang, Z. et al. 2006. Toward a Phylogenetic Classification of the Leotiomycetes Based on rDNA Data. *Mycologia*, 98(6): 1065 - 1075.
- Ward, N.L. et al. 2009. Three Genomes from the Phylum *Acidobacteria* Provide Insight into the Lifestyles of These Microorganisms in Soils. *Applied and Environmental Microbiology*, 75(7): 2046-2056.
- Wardle, D. A., 2002. *Communities and Ecosystems: Linking the Aboveground and Belowground Components*. New Jersey: Princeton University Press.
- Wardle, D. A. & Yeates, G. W., 1993. The Dual Importance of Competition and Predation as Regulatory Forces in Terrestrial Ecosystems: Evidence from Decomposer Food-Webs. *Oecologia*, 93(2): 303-306.
- Westover, K. M., Kennedy, A. C. & Kelley, S. E., 1997. Patterns of Rhizosphere Microbial Community Structure Associated with Co-Correlating Plant Species. *Journal of Ecology*, 85: 863-873.
- Whitehead, D. C., Dibb, H. & Hartley, R. D., 1981. Extractant pH and the Release of Phenolic Compounds from Soils, Plant Roots and Leaf Litter. *Soil Biology and Biochemistry*, 13(5): 343-348.
- Whitman, W. B., Coleman, D. C. & Wiebe, D. C., 1998. Prokaryotes: The Unseen Majority. *PNAS*, 95:6578 - 6583.
- Wickham, H., 2009. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.
- Widjojatmodjo, M. N., Fluit, A. C. & Verhoef, J., 1994. Rapid Identification of Bacteria by PCR-Single-Strand Conformation Polymorphism. *Journal of Clinical Microbiology*, 32(12): 3002-3007.
- Wieland, G., Neumann, R. & Backhaus, H., 2001. Variation of Microbial Communities in Soil, Rhizosphere, and Rhizoplane in Response to Crop Species, Soil Type, and Crop Development. *Applied and Environmental Microbiology*, 67(12): 5849-5854.
- Willig, M. R., Kaufman, D. M. & Stevens, R. D., 2003. Latitudinal Gradients of Biodiversity. *Annual Review of Ecology, Evolution and Systematics*, 34:273-309.

- Wirth, S., 1999. Regional-Scale analysis of Soil Microbial Biomass and Soil Basal CO₂-Respiration in Northeastern Germany. *Sustaining the Global Farm*: 448-495.
- Wit, R. & Bouvier, T., 2006. "Everything is everywhere, but, the environment selects"; what did Baas Becking and Beijerinck really say?. *Environmental Microbiology*, 8(4): 755 - 758.
- Woese, C., Kandler, O. & Wheelis, M., 1990. Towards a Natural System of Organisms: Proposal for the domains Archaea, Bacteria and Eucarya. *PNAS*, 87: 4576-4579.
- Wolfe, B.E., Mummey, D.L., Rilling, M.C., Klironomos, N.J., 2007. Small-Scale Spatial Heterogeneity of Arbuscular Mycorrhizal Fungal Abundance and Community Composition in a Wetland Plant Community. *Mycorrhiza*, 17:175-183.
- Wolters, V., 2001. Biodiversity of Soil Animals and its Function. *European Journal of Soil Biology*, 37(4): 221 - 227.
- Wolters, V. et al. 2000. Effects of Global Changes on Above- and Belowground Biodiversity in Terrestrial Ecosystems: Implications for Ecosystem Functioning. *BioScience*, 50(12): 1089 - 1098.
- Wood, M., 1995. A Mechanism of Aluminium Toxicity to Soil Bacteria and Possible Ecological Implications. *Plant and Soil*, 171: 63-69.
- Wyld, H. C. & Partridge, E. H., 1975. Webster Universal Dictionary. International Edition ed. New York: Harver Educational Services Inc.
- Yeates, C. et al. 1998. Methods for Microbial DNA Extraction from Soil for PCR Amplification. *Biological Procedures Online*, 1(1): 40-47.
- Yin, B., Crowley, D., Sparovek, G., Melo, W., Bornman, J., 2000. Bacterial Functional Redundancy Along a Soil Reclamation Gradient. *Applied and Environmental Microbiology*, 66(10): 4361 – 4365.
- Young, C. C. et al. 1993. Polyvinylpyrrolidone-Agarose Gel Electrophoresis Purification of Polymerase Chain Reaction-Amplifiable DNA from Soils. *Applied and Environmental Microbiology*, 59(6): 1972-1974.
- Young, I. M. & Ritz, K., 2000. Tillage, habitat space and function of soil microbes. *Soil and Tillage Research*, 52:201-213.
- Yu, Z. & Morrison, M., 2004. Comparisons of Different Hypervariable Regions of rrs Genes for Use in Fingerprinting of Microbial Communities by PCR-Denaturing Gradient Gel Electrophoresis. *Applied and Environmental Microbiology*, 70(8): 4800 - 4806.

- Zellerhoff, N., Himmelbach, A., Dong, W. & Bieri, S., 2010. Nonhost Resistance of Barley to Different Fungal Pathogens is Associated with Largely Distinct, Quantitative Transcriptional Responses. *Plant Physiology*, 152(4): 2053-2066.
- Zhang, L. & Xu, Z., 2008. Assessing Bacterial Diversity in Soil. *Journal of Soils and Sediments*, 8(6): 379-388.
- Zhou, J., Bruns, M. A. & Tiedje, J. M., 1996. DNA Recovery from Soils of Diverse Composition. *Applied and Environmental Microbiology*, 62(2): 316:322.
- Zhou, J. Z. et al. 1997. Phylogenetic Diversity of a Bacterial Community Determined from Siberian Tundra Soil DNA. *Microbiology*, 143:3913-3919.
- Zinger, L., Coissac, E., Choler, P. & Geremia, R., 2009. Assessment of Microbial Communities by Graph Partitioning in a Study of Soil Fungi in Two Alpine Meadows. *Applied and Environmental Microbiology*, 75(18): 5863 - 5870.
- Zoetendal, E. G. et al. 2001. DNA Isolation Protocols Affect the Detection Limit of PCR Approaches of Bacteria in Samples from the Human Gastrointestinal Tract. *Systematic and Applied Microbiology*, 24(3): 405-410.